UNIVERSITÀ DEGLI STUDI
DI MILANO

DIPARTIMENTO DI SCIENZE DELL'INFORMAZIONE

Rapporto interno N. 321-08

Introduzione alle catene di Markov finite

Massimiliano Goldwurm

Introduzione alle catene di Markov finite

Massimiliano Goldwurm

Rapporto Interno
RI 321-08

Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
via Comelico 39/41
20135 Milano, Italy

Maggio 2008

In questo lavoro si illustrano le principali proprietà classiche delle catene di Markov finite. La presentazione ha un carattere didattico e divulgativo. Si intende in particolare fornire agli studenti materiale didattico di supporto per il corso sui modelli stocastici Markoviani, tenuto a partire dall'anno accademico 2006/2007 nell'ambito dei corsi di laurea magistrale di area Informatica.

Indice

Capitolo 1

Matrici non negative

In questo capitolo introduciamo le proprietà principali delle matrici a coefficienti reali non negativi. L'obiettivo principale è quello di presentare il teorema di Perron-Frobenius sull'autovalore principale delle matrici primitive e alcune sue conseguenze riguardanti l'espressione asintotica delle potenze di tali matrici. Questi risultati sono legati allo studio delle proprietà ergodiche delle catene di Markov e saranno riconsiderati nei capitoli successivi. Ogni matrice non negativa può essere associata in modo del tutto naturale a un grafo orientato nel quale i lati sono pesati con valori positivi. Questa associazione permette di visualizzare in maniera semplice e diretta le proprietà di periodicità che in realtà dipendono direttamente dai cammini del grafo considerato. Per questo motivo iniziamo questo studio partendo proprio dai grafi orientati.

1.1 Grafi orientati

Ricordiamo che un *grafo orientato* è una coppia (V,E), dove V è un insieme finito di elementi detti nodi o vertici ed E un insieme di coppie ordinate di nodi, che chiamiamo lati o archi. In questa sede rappresenteremo sempre i nodi mediante interi positivi, quindi $V = \{1,2,\dots,k\}$ per un qualche $k \in \mathbb{N}$; di conseguenza un lato è una coppia (i,j), dove $i,j \in \{1,2,\dots,k\}$.

Come al solito rappresentiamo un grafo orientato mediante un diagramma nel quale i nodi sono punti o cerchi e i lati sono frecce che connettono tali punti. La figura 1.1 definisce un grafo orientato di 12 nodi.

In un grafo orientato G, un *cammino* è una sequenza di nodi $C = (i_0, i_1, \dots, i_n)$, non necessariamente

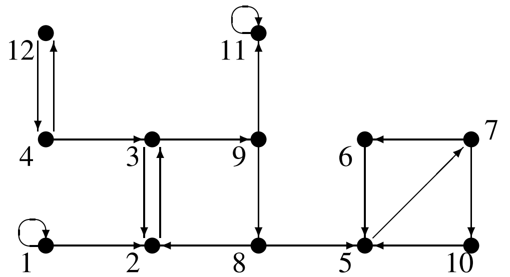


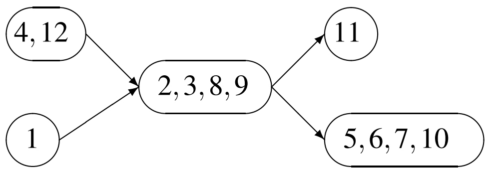
Figura 1.1: Esempio di grafo orientato

distinti, tale che (i_{t-1}, i_t) è un lato di G per ogni $t = 1, 2, \ldots, n$. Diciamo inoltre che C è un cammino dal primo nodo i_0 all'ultimo nodo i_n e che la sua lunghezza è n (ovvero il numero dei suoi lati). I cammini di lunghezza 0 sono quelli formati da un solo nodo, della forma $C = (i_0)$, mentre quelli di lunghezza 1 coincidono con i lati. Il cammino C inoltre si dice *semplice* se tutti i suoi nodi sono distinti tranne al più il primo e l'ultimo. Un *ciclo* è un cammino (non necessariamente semplice) di lunghezza non nulla nel quale il primo e l'ultimo nodo coincidono. I cicli di lunghezza 1 sono anche detti cappi. Nel grafo della figura 1.1 ci sono due cappi, rispettivamente sui nodi 1 e 11, due cicli semplici di lunghezza 2, un ciclo semplice di lunghezza 4 e due di lunghezza 3. Nota che se per un nodo passa un ciclo C allora ne passano infiniti, tutti quelli che si ottengono ripetendo C un numero arbitrario di volte.

Dato un grafo orientato $G = (V, E)$ possiamo definire tra i suoi nodi la relazione di connessione $\to$, quella di comunicazione $\leftrightarrow$ e quella di comunicazione in n passi $\xrightarrow{n}$. Per ogni $i, j \in V$, poniamo $i \to j$ se esiste un cammino da i a j in G. Vale inoltre $i \leftrightarrow j$ se $i \to j$ e $j \to i$, ovvero se esiste un cammino da i a j e uno da j a i. Infine, per ogni $n \in \mathbb{N}$, poniamo $i \xrightarrow{n} j$ se esiste un cammino di lunghezza n da i a j.

Il grafo G si dice *fortemente connesso* se, per ogni coppia di nodi i, j, vale la relazione $i \leftrightarrow j$. In questo caso ogni nodo è connesso a qualunque altro nodo del grafo. Più in generale invece, è facile verificare che la relazione $\leftrightarrow$ è una relazione di equivalenza, soddisfa cioè le proprietà riflessiva, simmetrica e transitiva ([1]). Essa quindi suddivide l'insieme dei nodi di G in classi di equivalenza che chiamiamo semplicemente *classi* di nodi (o *classi irriducibili*). Il sottografo formato da una classe di nodi C e dai lati di G che connettono tali nodi è chiamata *componente fortemente connessa* di G. Per esempio nel grafo rappresentato in figura 1.1 le classi sono date dagli insiemi $\{1\}$, $\{2,3,8,9\}$, $\{4,12\}$, $\{5,6,7,10\}$, $\{11\}$.

Una classe A è minore di una classe B, in simboli $A \leq B$, se $a \to b$ per qualche $a \in A$ e qualche $b \in B$. Tale relazione $\leq$ è una relazione di ordine parziale definita sull'insieme delle classi di G. Essa può essere facilmente rappresentata dal cosiddetto *grafo ridotto* di G nel quale i vertici sono le classi di G e i lati sono le coppie di classi (A, B), $A \neq B$, tali che per qualche $a \in A$ e qualche $b \in B$, (a, b) è un lato di G. Chiaramente in un grafo ridotto non possono esservi cicli (si tratta quindi di un grafo diretto aciclico). Nell'esempio di figura 1.1 il grafo ridotto è rappresentato dalla seguente immagine.



Una classe di nodi C si dice *essenziale* se, per ogni $i \in C$, $i \to j$ implica $j \in C$. Quindi non è possibile uscire da una classe essenziale seguendo un cammino che parte da un suo nodo. Una classe essenziale è una classe massimale rispetto alla relazione $\leq$ definita sopra. Nota che per ogni classe non essenziale C e per ogni nodo $i \in C$, esiste $j \notin C$ tale che $i \to j$. Nell'esempio di figura 1.1 le classi essenziali sono $\{11\}$ e $\{5,6,7,10\}$. Analogamente, diciamo che un nodo è essenziale se appartiene a una classe essenziale. Questo significa che un nodo i è essenziale se e solo se per ogni nodo j, $i \to j$ implica $j \to i$.

[1] Avvisiamo che in molti testi la relazione di connessione $\to$ è definita in modo diverso, ponendo $i \to j$ se esiste un cammino da i a j di lunghezza non nulla; qui preferiamo utilizzare la definizione data proprio perché in questo modo manteniamo la riflessività.

1.1.1 Periodicità

Dato un grafo orientato G, consideriamo un nodo i per il quale passa almeno un ciclo; chiamiamo *periodo* di i, e lo denotiamo $d(i)$, il massimo comun divisore delle lunghezze dei cicli che passano per i:

$$d(i) = \text{mcd}\{n \in \mathbb{N} \mid i \xrightarrow{n} i\}$$

Per esempio, nel grafo di figura 1.1 i nodi 2, 3 e 4 hanno periodo 2, mentre i nodi 5 e 6 hanno periodo 3. Diciamo che un nodo è *aperiodico* se il suo periodo è 1. Nota che ogni nodo dotato di un cappio è aperiodico. In realtà si può provare che il periodo di un nodo dipende dalla sua classe, ovvero tutti i nodi di una data classe hanno lo stesso periodo. In altre parole la periodicità è una proprietà delle classi.

Proposizione 1.1 *Dato un grafo orientato G consideriamo due nodi distinti i, j tali che $i \leftrightarrow j$. Allora $d(i) = d(j)$.*

Dimostrazione. Sia C un qualunque ciclo che passa per j e sia s la sua lunghezza: proviamo innanzitutto che $d(i)$ è divisore di s. Infatti, poiché $i \leftrightarrow j$ esiste un cammino C_1 da i a j ed esiste un cammino C_2 da j a i; siano u e v le lunghezze di tali cammini. Allora C_1 e C_2 formano un ciclo di lunghezza $u+v$ passante per i e quindi $d(i)$ è un divisore di $u+v$; analogamente $d(i)$ è un divisore di $u+s+v$ perchè C_1, C e C_2 formano un ciclo per i. Ne segue che $d(i)$ divide anche la differenza $(u+s+v)-(u+v) = s$, e quindi $d(i)$ è divisore della lunghezza di qualsiasi ciclo passante per j; poiché $d(j)$ è il massimo di tali divisori otteniamo $d(i) \leq d(j)$. Ripetendo lo stesso ragionamento con i e j scambiati, otteniamo $d(j) \leq d(i)$ e questo prova l'asserto. $\square$

Chiamiamo quindi periodo di una classe il periodo di uno dei suoi nodi (e quindi di tutti). Nel grafo di figura 1.1 le classi $\{1\}$ e $\{11\}$ hanno periodo 1, le classi $\{4,12\}$ e $\{2,3,8,9\}$ hanno periodo 2, mentre $\{5,6,7,10\}$ ha periodo 3.

1.2 Matrici e grafi

Ricordiamo anzitutto la notazione tradizionale sulle matrici. Una matrice reale $A = [a_{ij}]$ si dice *positiva*, in simboli $A > 0$, se $a_{ij} > 0$ per ogni i, j; A si dice invece *non negativa*, rispettivamente $A \geq 0$, se $a_{ij} \geq 0$ per tutte le coppie di indici. Più in generale, se $A = [a_{ij}]$ e $B = [b_{ij}]$ sono matrici reali di ugual dimensione, l'espressione $A \geq B$ significa che $a_{ij} \geq b_{ij}$ per ogni coppia di indici i e j. Analogamente si definiscono le relazioni $A > B$, $A \leq B$, $A < B$, $A = B$.

Consideriamo una matrice quadrata $A = [a_{ij}]$ di ordine k a coefficienti reali e supponiamo $A \geq 0$. Sia $G = (V,E)$ il grafo orientato tale che $V = \{1,2,\ldots,k\}$ ed $E = \{(i,j) \in V \times V \mid a_{ij} > 0\}$. Diciamo che G è il grafo *associato* alla matrice A e che a_{ij} è il peso del lato (i,j) per ogni $(i,j) \in E$. In questo modo A può essere considerata come una matrice di pesi positivi sui lati di G. Se inoltre i coefficienti di A assumono solo valori in $\{0,1\}$ (ovvero tutti i lati hanno peso 1), diciamo che A è la *matrice di adiacenza* di G. Per esempio, la matrice di adiacenza del grafo di figura 1.1 è la seguente:

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	0	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	1	0	0	0
4	0	0	1	0	0	0	0	0	0	0	0	1
5	0	0	0	0	0	1	1	0	0	0	0	0
6	0	0	0	0	0	0	1	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	1	0	0
8	0	1	0	0	1	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	0	0	1	0
10	0	0	0	0	1	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	1	0
12	0	0	0	1	0	0	0	0	0	0	0	0

(1.1)

La nozione di peso può essere estesa a tutti i cammini di G di lunghezza maggiore di 0. Dato un cammino $S = (i_0, i_1, \ldots, i_n)$ in G, dove $n \geq 1$, chiamiamo *peso* di S il prodotto

$$a_{i_0 i_1} a_{i_1 i_2} \cdots a_{i_{n-1} i_n}.$$

Tali valori sono sempre positivi. Inoltre, se A è la matrice di adiacenza di G allora tutti i cammini hanno peso 1.

Vogliamo ora mettere in rilievo la relazione esistente tra le potenze di A e i pesi dei cammini di data lunghezza. A tale scopo, per ogni $n \in \mathbb{N}$, denotiamo con $a_{ij}^{(n)}$ i coefficienti di A^n, per cui $A^n = [a_{ij}^{(n)}]$. Se $n \geq 2$ è facile verificare che ogni $a_{ij}^{(n)}$ è dato da

$$a_{ij}^{(n)} = \sum_{\ell=1}^{k} a_{i\ell}^{(n-1)} a_{\ell j} = \sum_{\substack{i_0 = i, i_n = j \\ i_t \in V \forall t = 1, \ldots, n-1}} a_{i_0 i_1} a_{i_1 i_2} \cdots a_{i_{n-1} i_n} \tag{1.2}$$

dove la seconda somma è estesa a tutte le sequenze di $n+1$ indici $(i_0, i_1, i_2, \ldots, i_n)$ tali che $i_0 = i$ e $i_n = j$.

Proposizione 1.2 *Sia $A = [a_{ij}]$ una matrice quadrata non negativa e sia G il grafo associato. Allora, per ogni intero positivo n e ogni coppia di indici i, j, il coefficiente $a_{ij}^{(n)}$ della matrice A^n rappresenta la somma dei pesi di tutti i cammini di lunghezza n in G dal nodo i al nodo j.*

Dimostrazione. Ragioniamo per induzione su n. Per $n = 1$ la proprietà è ovvia. Supponiamo la proprietà vera per un valore $n \geq 1$ e dimostriamola vera per $n+1$. Sia k la dimensione di A. Per definizione di prodotto di matrici, sappiamo che ogni coefficiente di A^{n+1} è dato da

$$a_{ij}^{(n+1)} = \sum_{r=1}^{k} a_{ir}^{(n)} a_{rj} \tag{1.3}$$

Per ipotesi di induzione, $a_{ir}^{(n)}$ è la somma dei pesi dei cammini di lunghezza n da i a r e quindi il prodotto $a_{ir}^{(n)} a_{rj}$ è la somma dei pesi dei cammini di lunghezza $n+1$ da i a j che passano da r prima di giungere in j. Sommando tali prodotti $a_{ir}^{(n)} a_{rj}$ al variare r in tutti i modi possibili otteniamo proprio la somma di tutti i cammini di lunghezza $n+1$ da i a j; di conseguenza la proprietà segue dalla relazione (1.3). □

Una conseguenza immediata della proposizione precedente è che, per ogni i, j e n, $a_{ij}^{(n)} > 0$ se e solo se in G esiste un cammino di lunghezza n da i a j.

Inoltre, se A è la matrice di adiacenza di G, allora la relazione (1.2) determina proprio il numero di cammini di lunghezza n che vanno da i a j.

Corollario 1.3 *Se $A = [a_{ij}]$ è la matrice di adiacenza di un grafo orientato G allora, per ogni intero positivo n e ogni coppia di indici i, j, il coefficiente $a_{ij}^{(n)}$ rappresenta il numero di cammini di lunghezza n in G dal nodo i al nodo j.*

1.3 Decomposizione di matrici

Una matrice quadrata non negativa $A = [a_{ij}]$ si dice *irriducibile* se per ogni coppia di indici i, j, esiste un intero $n > 0$, in generale dipendente da i e da j, tale che $a_{ij}^{(n)} > 0$. Questo significa che nel grafo associato ad A, per ogni coppia di nodi i, j esiste un intero $n > 0$ tale che $i \xrightarrow{n} j$ (nota che n non può essere nullo).

Quindi, il grafo associato a una matrice irriducibile è fortemente connesso; inoltre, per la proposizione 1.1, tutti i suoi nodi hanno lo stesso periodo. Definiamo allora il *periodo* di una matrice irriducibile come il periodo d di uno dei nodi del grafo associato. Chiaramente d non dipende dal nodo scelto. Diciamo allora che la matrice è *periodica* se $d > 1$; se invece $d = 1$ la matrice è detta *aperiodica*.

Osserva che l'unico grafo fortemente connesso associato a una matrice non irriducibile è quello formato da un solo nodo e privo di lati. In questo caso la matrice associata è banalmente la matrice 0. Ogni altro grafo fortemente connesso possiede matrice di adiacenza irriducibile.

Una matrice quadrata non negativa A, diversa da 0, si dice *decomponibile* (o anche riducibile) se non è irriducibile. Questo significa che una matrice è decomponibile se e solo se il grafo associato contiene almeno due componenti fortemente connesse.

Le matrici decomponibili possono essere poste in una forma particolare, detta triangolare a blocchi, permutando opportunamente l'ordine delle righe e delle colonne. Infatti, data una matrice decomponibile A, supponiamo che il grafo associato possieda r componenti fortemente connesse. Poiché $r > 1$, possiamo scambiare tra loro alcune righe e le corrispondenti colonne in modo tale che siano verificate le due condizioni seguenti:

1. per ogni classe D le righe (e le colonne) appartenenti a D sono poste consecutivamente,
2. le righe (e le colonne) di ciascuna classe D precedono quelle di un'altra classe F ogniqualvolta $D \leq F$ (cioé vi sia un lato da un nodo in D a uno in F).

Otteniamo così una matrice della forma

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1r} \\ [0] & A_{22} & A_{23} & \cdots & A_{2r} \\ [0] & [0] & A_{33} & \cdots & A_{3r} \\ & & \cdots & & \\ [0] & [0] & [0] & \cdots & A_{rr} \end{bmatrix} \tag{1.4}$$

dove le matrici A_{ii}, $i = 1, 2, \ldots, r$, si trovano sulla diagonale principale e rappresentano le matrici dei pesi delle r componenti fortemente connesse, mentre le matrici [0] sono a componenti nulle. Nota che le matrici A_{ii} diverse dalla matrice 0, sono irriducibili; invece, le matrici A_{ij} con $i < j$ rappresentano i pesi dei lati che connettono nodi della componente i-esima con nodi della componente j-esima. Le matrici della forma (1.4) sono dette *triangolari a blocchi*. Chiaramente una matrice quadrata non negativa è decomponibile se e solo se mediante una permutazione di righe e delle corrispondenti colonne può essere trasformata in una matrice triangolare a blocchi.

Per esempio, nel caso della matrice (1.1), associata al grafo in figura 1.1, otteniamo la seguente forma triangolare a blocchi.

$$
\begin{array}{c|cccccccccccc}
 & 1 & 4 & 12 & 2 & 3 & 8 & 9 & 5 & 6 & 7 & 10 & 11 \\
\hline
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
12 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
8 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
9 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
11 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}
\tag{1.5}
$$

Osserviamo che la forma triangolare a blocchi di una matrice decomponibile A è data da un prodotto del tipo PAP', dove P è una matrice di permutazione. Ricordiamo che una matrice di permutazione è una matrice a componenti 0 e 1 ottenuta dalla matrice identità permutando l'ordine delle righe e delle colonne (in modo corrispondente). È noto che ogni matrice di permutazione P è ortogonale, ovvero la sua trasposta P' coincide con la sua inversa ($PP' = I$).

1.4 Matrici irriducibili

Nella sezione precedente abbiamo introdotto le matrici irriducibili e abbiamo visto che ogni matrice quadrata non negativa può essere essenzialmente trasformata in una matrice triangolare a blocchi che possiede sulla diagonale principale matrici irriducibili (o la matrice 0). Vediamo ora alcune proprietà importanti delle matrici irriducibili, riguardanti le lunghezze dei cicli e dei cammini che connettono i vari nodi del grafo associato. Si tratta principalmente di risultati sulla periodicità di tali lunghezze, basati su proprietà classiche del massimo comun divisore di interi ricordate nella sezione A.4.

Nelle seguenti proposizioni supponiamo che $A = [a_{ij}]$ sia una matrice irriducibile di dimensione k e che d sia il suo periodo.

Proposizione 1.4 *Per ogni indice i,* $1 \leq i \leq k$*, esiste un intero* $N(i) > 0$ *tale che*

$$a_{ii}^{(nd)} > 0$$

per ogni intero $n \geq N(i)$.

Dimostrazione. Sia $L(i)$ l'insieme $\{t \in \mathbb{N} \mid a_{ii}^{(t)} > 0\}$. Nota che $L(i)$ è la famiglia delle lunghezze dei cicli che passano per i nel grafo associato ad A e quindi d è il suo massimo comun divisore. Si verifica che $L(i)$ è chiuso rispetto alla somma: infatti, se $t_1, t_2 \in L(i)$ allora

$$a_{ii}^{(t_1+t_2)} \geq a_{ii}^{(t_1)} a_{ii}^{(t_2)} > 0$$

e quindi anche $t_1+t_2 \in L(i)$. Di conseguenza, per il teorema A.4, $L(i)$ contiene tutti i multipli di d a partire da un certo intero in poi e questo prova il risultato. $\square$

Proposizione 1.5 *Fissato un indice i arbitrario, per ogni indice j esistono due interi r_j e q_j, $0 \le r_j < d$, $0 < q_j$, tali che*

a) *se $a_{ij}^{(s)} > 0$ allora $s \equiv r_j \pmod d$;*

b) *$a_{ij}^{(r_j+nd)} > 0$ per ogni $n \ge q_j$.*

Dimostrazione. Siano s_1 e s_2 tali che $a_{ij}^{(s_1)} > 0$ e $a_{ij}^{(s_2)} > 0$; poiché $j \to i$, esiste $t \in N$ tale che $a_{ji}^{(t)} > 0$. Ne segue che s_1+t e s_2+t sono lunghezze di cicli passanti per i e quindi d è un loro divisore. Di conseguenza d divide anche la loro differenza, ovvero $s_1+t-(s_2+t) = s_1 - s_2 \equiv 0 \pmod d$ e questo implica $s_1 \equiv s_2 \pmod d$. In questo modo abbiamo provato la proprietà a).

Sia ora s tale $a_{ij}^{(s)} > 0$. Per la a) sappiamo che $s = r_j + n_0 d$, per qualche $n_0 \in \mathbb{N}$. Inoltre, per la proposizione 1.4, esiste $N(j)$ tale che $a_{jj}^{(nd)} > 0$, per ogni $n > N(j)$. Di conseguenza, posto $q_j = n_0 + N(j)$ e preso un intero $n \ge q_j$, abbiamo

$$a_{ij}^{(r_j+nd)} \ge a_{ij}^{(r_j+n_0 d)} a_{jj}^{((n-n_0)d)} > 0$$

e questo implica la proprietà b). $\square$

La proposizione precedente consente di raggruppare gli indici j a seconda della classe di resto r_j determinata dalla proprietà a). Più precisamente, fissato un indice iniziale i, per ogni $r \in \{0,1,\dots,d-1\}$ consideriamo l'insieme C_r degli indici j tali che $r_j = r$. Si verifica che $C_r \neq \emptyset$ per ogni r e che l'unione $\bigcup_{r=0}^{d-1} C_r$ coincide con l'insieme di tutti i possibili indici $\{1,2,\dots,k\}$. La famiglia $\{C_r \mid r = 0,1,\dots,d-1\}$ forma così una partizione di $\{1,2,\dots,k\}$. Si può dimostrare che tale partizione non dipende dall'indice iniziale i. Cambiando l'indice iniziale si ottiene la stessa famiglia di sottoinsiemi C_r a meno di una permutazione ciclica dei coefficienti r.

Si può verificare facilmente quest'ultima proprietà. Infatti, sia r_j la classe di resto corrispondente all'indice j rispetto all'indice iniziale i e sia r'_j lo stesso valore assumendo l'indice iniziale $i' \neq i$. Dalla proposizione precedente deduciamo che per s e t abbastanza grandi, abbiamo

$$a_{ij}^{(r_{i'}+sd+r'_j+td)} \ge a_{ii'}^{(r_{i'}+sd)} a_{i'j}^{(r'_j+td)} > 0$$

e quindi per la proprietà a)

$$r_{i'} + sd + r'_j + td \equiv r_j \pmod d$$

ovvero $r_{i'} + r'_j \equiv r_j \pmod d$. Questo significa appunto che tutti i valori r'_j sono ottenuti dai valori r_j attraverso una permutazione ciclica determinata dalla lunghezza $r_{i'}$.

Usando la partizione $\{C_r \mid r = 0,1,\dots,d-1\}$, possiamo allora permutare le righe e le corrispondenti colonne trasfomando la matrice in una forma canonica nella quale tutti gli indici nella stessa classe sono posti consecutivamente e le classi C_r sono disposte in ordine crescente rispetto a r. La matrice ottenuta è del tipo

$$A' = \begin{bmatrix} [0] & A_0 & [0] & \cdots & [0] \\ [0] & [0] & A_1 & \cdots & [0] \\ [0] & [0] & [0] & \cdots & [0] \\ & & \cdots & & \\ [0] & [0] & [0] & \cdots & A_{d-2} \\ A_{d-1} & [0] & [0] & \cdots & [0] \end{bmatrix}$$

dove le sottomatrici [0] sulla diagonale principale sono quadrate e ogni A_i è la sottomatrice (non necessariamente quadrata) dei pesi dei lati da indici in C_i a indici in $C_{i+1(\text{mod } d)}$.

1.5 Matrici primitive

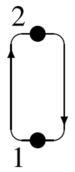
Le matrici primitive sono un tipo particolare di matrici irriducibili. Formalmente, una matrice non negativa A si dice *primitiva* se esiste un intero $n > 0$ tale che $A^n > 0$. Questo significa che nel grafo associato, per ogni coppia di nodi i e j abbiamo $i \xrightarrow{n} j$. Per esempio la matrice

$$B = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

è una matrice primitiva perchè $B^2 > 0$, ovvero tutte le coppie di nodi possono essere connesse da un cammino di lunghezza 2.

Dalla definizione segue immediatamente che una matrice primitiva è anche irriducibile. Viceversa, vi sono matrici irriducibili che non sono primitive. Per esempio, nella seguente figura la matrice D è irriducibile ma non primitiva; essa rappresenta il grafo G_1 ed è facile verificare che nessuna delle sue potenze è strettamente positiva.

$$D = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad G_1 =$$



In particolare, se m è pari abbiamo $D^m = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$; mentre se m è dispari si verifica $D^m = D$.

Inoltre, se A è una matrice primitiva e $A^n > 0$, allora $A^m > 0$ per ogni $m \geq n$: di conseguenza tutti gli indici di A sono aperiodici perchè possiedono un ciclo di lunghezza m per ogni $m \geq n$. Più in generale si può provare il seguente risultato.

Proposizione 1.6 *Una matrice non negativa A è primitiva se e solo se A è irriducibile e aperiodica.*

Dimostrazione. Se A è primitiva, per la discussione precedente sappiamo che A è irriducibile e aperiodica. Viceversa, supponiamo che A sia irriducibile e abbia periodo $d = 1$. Applichiamo la proposizione 1.5 alla matrice A. Chiaramente, tutti i coefficienti r_j in questo caso sono uguali a 0. Inoltre, per l'enunciato b), sappiamo che per ogni i, j esiste un intero $q(i, j) > 0$ tale che $a_{ij}^{(n)} > 0$ per ogni $n \geq q(i, j)$; scegliamo allora $N = \max\{q(i, j) \mid i, j\}$ e otteniamo che, per ogni $n \geq N$, $a_{ij}^{(n)} > 0$ per tutti gli indici i, j. Questo significa $A^n > 0$ e quindi A è primitiva. $\square$

1.6 Il teorema di Perron-Frobenius

La proprietà principale delle matrici primitive è data dal seguente teorema che qui riportiamo senza dimostrazione. In questa sede ci preme innanzitutto mettere in rilievo questa proprietà e ricordare alcune importanti conseguenze. La dimostrazione del teorema e delle sue principali conseguenze è rimandata nell'appendice al capitolo B. Per maggiori dettagli si può consultare [9].

Teorema 1.7 *(Teorema di Perron-Frobenius per matrici primitive) Sia $A \geq 0$ una matrice primitiva. Allora esiste un autovalore λ di A che gode delle seguenti proprietà:*

1. *λ è reale e maggiore di 0;*
2. *ogni autovalore μ di A diverso da λ verifica la relazione $|\mu| < \lambda$;*
3. *λ ammette autovettori destri e sinistri strettamente positivi;*
4. *λ è radice semplice dell'equazione caratteristica di A (la sua molteplicità algebrica è 1).*

La costante λ è anche chiamata autovalore di Perron-Frobenius della matrice A.

Diverse proprietà utili possono essere derivate dal teorema precedente. La principale riguarda l'ordine di grandezza delle potenze di matrici primitive. Essa asserisce che la potenza A^n di una matrice primitiva A cresce con ordine di grandezza uguale a λ^n, la differenza inoltre è determinata dal modulo massimo degli altri autovalori e dalla loro molteplicità.

Teorema 1.8 *Sia A una matrice primitiva, e siano $\lambda, \mu_1, \mu_2, \ldots, \mu_t$ i suoi autovalori distinti, disposti in modo tale che $\lambda > |\mu_1| \geq |\mu_2| \geq \ldots \geq |\mu_t|$ e la molteplicità m di μ_1 sia maggiore o uguale a quella degli autovalori aventi lo stesso modulo di μ_1. Allora, al crescere di n, abbiamo*

$$A^n = \lambda^n uv' + O(n^{m-1}|\mu_1|^n) \tag{1.6}$$

dove v' e u sono rispettivamente autovettore sinistro e autovettore destro corrispondenti a λ, normalizzati in modo tale che $v'u = 1$.

Nota che l'uguaglianza (1.6) può essere interpretata come una estensione della equazione (A.2) al caso delle matrici primitive (che in generale possono avere autovalori di molteplicità algebrica maggiore di 1).

Ricordiamo infine le seguenti proprietà la cui dimostrazione può essere trovata in [9], nelle quali $A = [a_{ij}]$ è una matrice primitiva e λ il suo autovalore di Perron-Frobenius.

1. λ soddisfa la relazione
$$\min_i \sum_j a_{ij} \leq \lambda \leq \max_i \sum_j a_{ij}$$
con l'uguaglianza che vale solo se i valori $\sum_j a_{ij}$ sono uguali per tutte le righe i (inoltre una proprietà analoga vale per le colonne);

2. se B è una matrice delle stesse dimensioni di A tale che $0 \leq B \leq A$ e β è un suo autovalore, allora $|\beta| \leq \lambda$ e inoltre $|\beta| = \lambda$ se e solo se $B = A$;

3. se v e u sono rispettivamente autovettore sinistro e autovettore destro di λ, normalizzati in modo tale che $v'u = 1$ (ovvero definiti come nel teorema 1.8), allora la matrice uv' è data da
$$uv' = \frac{Adj(\lambda I - A)}{\varphi'(\lambda)}$$
dove $\varphi(x) = \det(xI - A)$.

Concludiamo osservando che un analogo del teorema 1.7 vale anche per le matrici irriducibili; l'unica differenza è che nel caso di matrici periodiche esistono più autovalori di modulo massimo.

Teorema 1.9 *Se $A \geq 0$ è una matrice irriducibile allora esiste un autovalore λ di A che verifica le proprietà 1, 3, 4 del teorema 1.7 e tale che $|\mu| \leq \lambda$ per ogni autovalore μ di A. Inoltre, se il periodo di A è $d \geq 1$ allora le costanti $\lambda e^{i2s\pi/d}$, $s = 0, 1, \ldots, d-1$, sono gli autovalori di A di modulo λ.*

Nota che gli autovalori di modulo massimo di una matrice irriducibile di periodo d coincidono con le radici complesse dell'equazione $x^d - \lambda^d = 0$.

1.7 Matrici stocastiche

Finora abbiamo considerato generiche matrici a coefficienti non negativi. Tuttavia, nei capitoli successivi utilizzeremo matrici i cui coefficienti sono probabilità e risulterà naturale supporre che la somma dei valori disposti su ogni riga sia uguale a uno. Le matrici che godono di questa proprietà sono dette stocastiche.

Formalmente, un vettore riga $v = (v_1, v_2, \ldots, v_r) \in \mathbb{R}^r$ si dice *stocastico* se $0 \leq v_i \leq 1$ per ogni indice i e inoltre $\sum_{i=1}^r v_i = 1$ (e lo stesso dicasi per i vettori colonna). Analogamente, una matrice $P = [p_{ij}]$ si dice *stocastica* se ogni sua riga è un vettore stocastico, ovvero se $0 \leq p_{ij} \leq 1$ per ogni coppia di indici i, j, e inoltre $\sum_{j \geq 1} p_{ij} = 1$ per ogni i. Nel seguito supporremo sempre che una matrice stocastica sia quadrata. Inoltre una matrice doppiamente stocastica è una matrice stocastica P nella quale anche le colonne sono vettori stocastici, ovvero $\sum_{i \geq 1} p_{ij} = 1$ per ogni j. È facile verificare che il prodotto di due matrici stocastiche è ancora una matrice stocastica. Valgono inoltre le seguenti proposizioni.

Proposizione 1.10 *Sia P una matrice stocastica di dimensione r. Allora 1 è autovalore di P e ammette come corrispondente autovettore destro il vettore e di dimensione r tale che $e' = (1, 1, \ldots, 1)$.*

Dimostrazione. Sia $P = [p_{ij}]$. Per la stessa definizione di matrice stocastica abbiamo, per ogni $i = 1, \ldots, r$, $(Pe)_i = \sum_j p_{ij} = 1$ e quindi vale

$$Pe = e$$

Questo prova che 1 è autovalore di P e che e è un autovettore destro di P corrispondente a 1. □

Per lo stesso motivo, se P è una matrice doppiamente stocastica allora $e'P = e'$; di conseguenza, e è anche autovettore sinistro di P corrispondente a 1.

Proposizione 1.11 *Se λ è un autovalore di una matrice stocastica allora $|\lambda| \leq 1$.*

Dimostrazione. Sia $P = [p_{ij}]$ una matrice stocastica di dimensione r e sia λ un suo autovalore. Consideriamo un autovettore sinistro v di P corrispondente a λ. Allora $v'P = \lambda v'$ e quindi abbiamo

$$\sum_{j=1}^r |\lambda v_j| = \sum_{j=1}^r |\sum_{i=1}^r p_{ij} v_i| \leq \sum_{i=1}^r |v_i| \, |\sum_{j=1}^r p_{ij}| = \sum_{i=1}^r |v_i|$$

Di conseguenza $|\lambda| \sum_{j=1}^r |v_j| \leq \sum_{i=1}^r |v_i|$ e questo implica $|\lambda| \leq 1$. □

Nel caso di matrici stocastiche primitive possiamo applicare i teoremi 1.7 e 1.8 alle proprietà precendenti ottenendo il seguente risultato.

Proposizione 1.12 *Se P è una matrice stocastica primitiva allora 1 è il suo autovalore di Perron-Frobenius e inoltre, per qualche $0 \leq \varepsilon < 1$, abbiamo*

$$P^n = ev' + O(\varepsilon^n)$$

dove v è l'autovettore sinistro di P corrispondente a 1 tale che $\sum_{i \geq 1} v_i = 1$.

Quindi, nella proposizione precedente v è un vettore stocastico. Inoltre ev' è una matrice *stabile*, cioè possiede tutte le righe uguali, in questo caso coincidenti con v'. Nota anche che, per $n \to +\infty$, P^n converge alla matrice ev' e la velocità di convergenza è esponenziale.

Nel caso di matrici doppiamente stocastiche e è anche autovettore sinistro corrispondente a 1 e quindi otteniamo il seguente enunciato.

Corollario 1.13 *Se P è una matrice primitiva doppiamente stocastica di dimensione r, allora l'autovettore sinistro v definito nella proposizione precedente è dato da $v' = (1/r, 1/r, \dots, 1/r)$.*

Capitolo 2

Nozioni fondamentali sulle catene di Markov

La nozione di catena di Markov fu implicitamente introdotta dal matematico russo A.A. Markov (1856-1922) in una serie di articoli, il primo dei quali apparso nel 1906, dedicati allo studio di sequenze di variabili aleatorie debolmente dipendenti. Lo scopo era quello di estendere a questo caso le note proprietà asintotiche valide nell'ipotesi di indipendenza delle variabili. La nozione di dipendenza qui considerata è relativamente debole, nel senso che nella sequenza ogni variabile dipende completamente dalla precedente. Per questo motivo essa è del tutto naturale e si applica a numerose situazioni. Negli anni '30 e '40 le catene di Markov sono state studiate e approfondite in maniera sistematica da diversi autori e oggi esiste una vasta letteratura sull'argomento, con innumerevoli applicazioni che riguardano non solo il calcolo delle probabilità ma anche le scienze naturali, quelle economiche, sociali e svariati altri settori. Per una bibliografia aggiornata agli anni '80 si può consultare [4], dove si possono trovare i riferimenti alle collezioni dei lavori originali di Markov e agli articoli dei principali autori successivi. Altri riferimenti più recenti si trovano per esempio in [11, 1, 3].

In questa sede ci occuperemo principalmente delle catene di Markov finite e omogenee. Prima di dare la definizione formale illustriamo la nozione attraverso alcuni classici esempi tratti da vari testi che si possono consultare per ulteriori informazioni [2, 6, 4, 11, 3].

2.1 Esempi

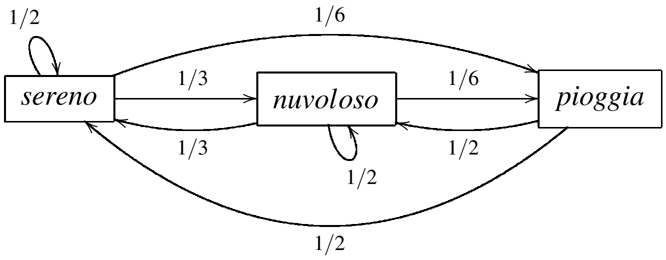
Un modello per il tempo atmosferico [6, 11]

Vogliamo definire un modello probabilistico per descrivere l'evoluzione del tempo in un periodo di riscaldamento climatico, caratterizzato da scarse precipitazioni. Per semplicità, assumiamo tre possibili condizioni del tempo in una giornata qualsiasi: sereno, nuvoloso, pioggia. Supponiamo che le condizioni di domani dipendano sempre dal tempo di oggi ma non da quello dei giorni precedenti. Inoltre, assumiamo che non vi siano mai due giornate di pioggia consecutive; se un giorno piove, la giornata successiva è serena o nuvolosa con uguale probabilità. Invece, dopo una giornata serena o nuvolosa il tempo rimane invariato con probabilità $1/2$; se si verifica un cambiamento allora piove in un terzo dei casi.

In queste ipotesi, le probabilità di transizione tra le varie condizioni atmosferiche in due giorni consecutivi possono essere rappresentate dalla seguente matrice:

	sereno	nuvoloso	pioggia
sereno	$1/2$	$1/3$	$1/6$
nuvoloso	$1/3$	$1/2$	$1/6$
pioggia	$1/2$	$1/2$	0

oppure dal grafo pesato riportato di seguito.



Domande naturali che possiamo porci sono le seguenti: se oggi è sereno qual è la probabilità che piova almeno una volta nei prossimi 5 giorni? Mediamente, quante volte piove nell'arco di un mese? Quanto dura in media un periodo di siccità?

La rovina del giocatore [2, 4].

Consideriamo un gioco nel quale due concorrenti, chiamati A e B rispettivamente, si disputano un capitale complessivo di r monete in una successione di partite consecutive. Inizialmente, A possiede k monete, mentre B ne possiede $r-k$, per un dato $k \in \{1,2,\dots,r-1\}$. A ogni partita A vince con una probabilità p prefissata, dove $0 < p < 1$, mentre B vince con probabilità $q = 1-p$. Inoltre, ogni partita è indipendente dalle altre e alla sua conclusione il perdente paga una moneta al vincitore. Il gioco termina quando uno dei due contendenti rimane senza monete.

Questo gioco può essere rappresentato da un insieme di stati $S = \{0,1,2,\dots,r\}$, dove lo stato corrente rappresenta il capitale di A. Quindi, k è lo stato iniziale del gioco, lo stato 0 rappresenta la rovina di A e lo stato r la rovina di B. Ogni partita determina la transizione dallo stato corrente i a uno dei due stati adiacenti $i+1$ o $i-1$, a seconda della vittoria di A o di B. Le probabilità di transizione tra i vari stati sono allora definite dalla matrice $P = [P(i,j)]_{i,j=0,1,\dots,r}$ tale che

$$P(i,i+1) = \begin{cases} p & \text{se } i \neq 0,r \\ 0 & \text{altrimenti} \end{cases}, \qquad P(i,i-1) = \begin{cases} q & \text{se } i \neq 0,r \\ 0 & \text{altrimenti} \end{cases}, \qquad P(0,0) = P(r,r) = 1 \ ,$$

mentre in tutti gli altri casi $P(i,j) = 0$. La matrice P è quindi la seguente:

	0	1	2	3	...	r-2	r-1	r
0	1	0	0	0	...	0	0	0
1	q	0	p	0	...	0	0	0
2	0	q	0	p	...	0	0	0
.	.	.	.	.	...	.	.	.
.	.	.	.	.	...	.	.	.
.	.	.	.	.	...	.	.	.
r-1	0	0	0	0	...	q	0	p
r	0	0	0	0	...	0	0	1

Lo schema precedente può rappresentare diversi altri processi aleatori. Uno di questi è la nota "passeggiata dell'ubriaco" (vedi per esempio [11]). Sono inoltre state considerate varie estensioni. Possiamo ad esempio sostituire i due stati assorbenti 0 e r con due barriere elastiche o riflettenti. Nel primo caso avremmo $P(0,0) = a$, $P(0,1) = b$, $P(r,r) = c$, $P(r,r-1) = d$, con $a,b,c,d \in (0,1)$, $a+b = 1$ e $c+d = 1$. Mentre nel secondo, $P(0,0) = P(r,r) = 0$ e $P(0,1) = P(r-1,r) = 1$.

Moto rettilineo di una particella

Molti esempi rilevanti possono essere rappresentati dal moto di una particella su un segmento di interi [2, 4]. Questo modello è definito da una particella che si muove sui punti $\{0,1,\ldots,r\}$, per un intero $r > 0$ fissato. Ad ogni istante, dalla posizione corrente i la particella si muove di una posizione avanti o indietro con probabilità rispettive p_i e q_i, oppure rimane ferma con probabilità m_i, dove $p_i, q_i, m_i \in [0,1]$ e $p_i + q_i + m_i = 1$ (chiaramente $q_0 = p_r = 0$). Le probabilità di transizione sono dunque rappresentate dalla seguente matrice:

	0	1	2	3	...	r-2	r-1	r
0	m_0	p_0	0	0	...	0	0	0
1	q_1	m_1	p_1	0	...	0	0	0
2	0	q_2	m_2	p_2	...	0	0	0
.	.	.	.	.	...	.	.	.
.	.	.	.	.	...	.	.	.
.	.	.	.	.	...	.	.	.
r-1	0	0	0	0	...	q_{r-1}	m_{r-1}	p_{r-1}
r	0	0	0	0	...	0	q_r	m_r

Passeggiate a caso in un grafo

Citiamo infine un altro esempio rilevante che riguarda il moto di un particella in un grafo. Dato un grafo non orientato $G = (V,E)$ supponiamo che una particella si muova tra i nodi di G partendo da un vertice prefissato; ad ogni passo si sceglie il nuovo vertice in modo equiprobabile tra tutti quelli adiacenti al nodo corrente. Per ogni $u \in V$ denotiamo con d_u il numero di vertici adiacenti a u; allora, le probabilità di transizione sono rappresentate dalla matrice $P = [P_{uv}]_{u,v\in V}$ dove, per ogni $u,v \in V$,

$$P_{uv} = \begin{cases} \frac{1}{d_u} & \text{se } \{u,v\} \in E \\ 0 & \text{altrimenti} \end{cases}$$

2.2 Definizione

Per definire formalmente una catena di Markov finita e omogenea abbiamo bisogno di tre quantità:

a) un insieme finito $S = \{1,2,\ldots,k\}$ di elementi, chiamati *stati*;

b) una distribuzione di probabilità μ definita su S che chiamiamo *distribuzione iniziale*, ovvero una funzione $\mu : S \longrightarrow \mathbb{R}$ tale che

$$\mu(i) \geq 0 \ \text{ per ogni } i \in S \quad \text{e inoltre} \quad \sum_{i\in S} \mu(i) = 1$$

c) una matrice stocastica P con indici in S, detta *matrice di transizione*, ovvero una famiglia di coefficienti $P = [p(i,j)]_{i,j\in S}$ tale che

$$p(i,j) \geq 0 \ \text{ per ogni } i,j \in S \ \text{ e inoltre } \sum_{j\in S} p(i,j) = 1 \ \text{ per ogni } i \in S$$

Una *catena di Markov finita e omogenea* con spazio degli stati S, distribuzione iniziale μ e matrice di transizione $P = [p(i, j)]$, è una sequenza $\{X_n\}_{n\in\mathbb{N}}$ di variabili aleatorie a valori in S tale che:

1. per ogni $i \in S$,
$$\Pr(X_0 = i) = \mu(i) \ ;$$

2. per ogni intero $n > 0$, ogni $i, j \in S$ e ogni n-pla di stati $i_0, i_1, \ldots, i_{n-1}$
$$\Pr(X_{n+1} = j \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i) = \Pr(X_{n+1} = j \mid X_n = i) \ ;$$

3. per ogni $n \in \mathbb{N}$ e ogni $i, j \in S$
$$\Pr(X_{n+1} = j \mid X_n = i) = p(i, j) \ .$$

La condizione 2 è nota come *proprietà di Markov*. Essa stabilisce che per ogni $n \in \mathbb{N}$, una volta noto il valore di X_n, l'evento $X_{n+1} = j$ non dipende dai valori assunti dalle variabili $X_0, X_1, \ldots, X_{n-1}$. In modo intuitivo, possiamo dire che la conoscenza del presente è sufficiente a determinare il futuro, nel senso che quest'ultimo non dipende dal passato.

La condizione 3 invece rappresenta l'omogeneità del processo; essa afferma che la probabilità di passare da uno stato i all'istante n-esimo a uno stato j all'istante $n+1$ non dipende da n ma solo da i e j. Tale probabilità rimane quindi la stessa per tutto il processo.

È naturale chiedersi se la definizione appena data descrive un'entità effettiva, che esiste davvero e che può essere costruita a partire dalle ipotesi fissate. In realtà si può provare che una tale sequenza $\{X_n\}$ esiste sempre. Più precisamente, per ogni tripla (S, μ, P) definita dalle condizioni **a)**, **b)**, **c)**, si può sempre costruire una catena di Markov corrispondente, cioè una sequenza di variabili aleatorie $\{X_n\}_{n\in\mathbb{N}}$ che soddisfi le proprietà 1, 2 e 3 date sopra.

Per costruire tale sequenza è necessario fissare lo spazio di probabilità sul quale tutte le X_n sono definite, a partire dalle quantità S, μ e P. A tal fine, chiamiamo *traiettoria* una qualunque sequenza di infiniti stati, ovvero un elemento $t = (t_0, t_1, \ldots, t_n \ldots)$, dove $t_n \in S$ per ogni $n \in \mathbb{N}$. Denotiamo inoltre con Ω l'insieme di tutte le traiettorie:

$$\Omega = \{t = (t_0, t_1, \ldots, t_n \ldots) \mid t_n \in S \ \text{ per ogni } \ n \in \mathbb{N}\}$$

e, per ogni $b \in S^{n+1}$ ($n \in \mathbb{N}$), $b = (b_0, b_1, \ldots, b_n)$, sia $C(b)$ il *cilindro* di base b, ovvero l'insieme di tutte le traiettorie che sono continuazioni di b:

$$C(b) = \{t \in \Omega \mid t_i = b_i \ \text{ per ogni } \ i = 0, 1, \ldots, n\}$$

Per convenzione consideriamo anche l'insieme Ω come un cilindro, ponendo $\Omega = C(\varepsilon)$, dove ε rappresenta il vettore vuoto. Inoltre definiamo la probabilità di un cilindro $C(b)$, con $b = (b_0, b_1, \ldots, b_n)$, ponendo
$$\Pr_\mu(C(b)) = \mu(b_0) p(b_0, b_1) p(b_1, b_2) \cdots p(b_{n-1}, b_n)$$
e fissando in particolare $\Pr_\mu(\Omega) = 1$.

Chiaramente l'insieme dei cilindri non forma una σ-algebra e di conseguenza la funzione $\Pr_\mu$ non è per ora una misura di probabilità. Tuttavia un noto risultato dovuto a Kolmogorov garantisce l'esistenza di un'unica minima σ-algebra $\mathcal{B}$ che include la famiglia di tutti i cilindri: a tale σ-algebra la funzione $\Pr_\mu$ può essere estesa in modo da costituire effettivamente una misura di probabilità. La costruzione della

σ-algebra $\mathcal{B}$ e la relativa estensione di Pr_μ sono argomenti classici di teoria della misura che esulano dagli scopi della nostra presentazione (per maggiori dettagli si veda [4, sez. 1.5 e 2.1] oppure [11, sez. 1.B]).

In questo modo, la tripla $(\Omega, \mathcal{B}, \mathrm{Pr}_\mu)$ forma uno spazio di probabilità sul quale possiamo definire le variabili aleatorie X_n, $n \in \mathbb{N}$, come semplici proiezioni, ponendo $X_n : \Omega \longrightarrow S$ tale che

$$X_n(t) = t_n \ , \qquad \text{per ogni } t \in \Omega, \, t = (t_0, t1, \ldots, t_k, \ldots) \ .$$

Si può ora verificare facilmente che la sequenza $\{X_n\}$ soddisfa le condizioni 1, 2 e 3 della definizione. In particolare, per ogni $n \in \mathbb{N}$, la probabilità congiunta delle variabili $X_0, X_1, \ldots, X_n$ è definita dalla probabilità dei corrispondenti cilindri: per ogni $b = (b_0, b_1, \ldots, b_n) \in S^{n+1}$,

$$\mathrm{Pr}_\mu(X_0 = b_0, \ldots, X_n = b_n) = \mathrm{Pr}_\mu(C(b)) = \mu(b_0)p(b_0, b_1)p(b_1, b_2) \cdots p(b_{n-1}, b_n)$$

Nel seguito supporremo sempre che ogni catena di Markov $\{X_n\}$ sia definita sullo spazio di probabilità $(\Omega, \mathcal{B}, \mathrm{Pr}_\mu)$ appena descritto (solitamente chiamato *spazio delle traiettorie*). Inoltre, se μ è puntuale, ovvero $\mu = \delta_i$ per qualche $i \in S$, allora rappresenteremo Pr_μ nella forma Pr_i. In particolare, per ogni distribuzione iniziale μ abbiamo $\mathrm{Pr}_i(X_n = j) = \mathrm{Pr}_\mu(X_n = j \mid X_0 = i)$. Denoteremo infine con E_i l'operatore di valor medio associato a Pr_i. Se per esempio Y è una variabile aleatoria a valori in $\mathbb{N}$ definita sullo spazio delle traiettorie, allora $E_i(Y) = \sum_{n \in \mathbb{N}} n \mathrm{Pr}_i(Y = n)$.

Osserviamo infine che ogni catena di Markov finita e omogenea, con insieme degli stati S e matrice di transizione $P = [p(i, j)]_{i,j \in S}$, può essere associata a un grafo orientato $G = (V, E)$, dove $V = S$, $E = \{(i, j) \mid p(i, j) > 0\}$ e ogni lato $(i, j) \in E$ è etichettato da $p(i, j)$. Diremo anche che G è il *grafo di connessione* della catena di Markov. In questo modo possiamo classificare gli stati della catena usando le relazioni di connessione $\rightarrow$ e di comunicazione $\leftrightarrow$ del grafo, come illustrato nel capitolo 1.

Nota che, per ogni $i, j \in S$, abbiamo $i \rightarrow j$ se $i = j$ oppure esiste una sequenza di stati $i_0, i_1, \ldots, i_n$ con $i_0 = i$ e $i_n = j$ tali che

$$\prod_{\ell=1}^{n} p(i_{\ell-1}, i_\ell) \ > \ 0$$

Questa condizione è equivalente a richiedere che $\mathrm{Pr}_i(X_n = j) > 0$ per qualche $n \in \mathbb{N}$.

Inoltre possiamo estendere in modo ovvio agli stati le nozioni e le proprietà introdotte per i nodi. Così una classe di stati non è altro che un insieme di nodi nel corrispondente grafo di connessione che forma una classe di equivalenza rispetto alla relazione $\leftrightarrow$. Analogamente parleremo di periodo di uno stato, di stati periodici e aperiodici e diremo che uno stato $i \in S$ è essenziale se i è un nodo essenziale nel grafo di connessione.

Una catena di Markov finita eredita anche in modo naturale le proprietà della matrice di transizione associata P. Così diremo che la catena è irriducibile se la matrice P è irriducibile. In modo analogo si definiscono le catene primitive, quelle periodiche e aperiodiche.

Osservazione 1 La definizione di catena di Markov che abbiamo introdotta può essere estesa in vari sensi. Se l'insieme degli stati S è infinito (ma numerabile) parleremo di catena di Markov *infinita*. Spesso nella letteratura la nozione di catena di Markov fa riferimento a questa definizione; nota che in questo caso la matrice $P = [p(i, j)]$ possiede infinite righe e infinite colonne.

Ricordiamo inoltre che una catena di Markov *non omogenea* è una sequenza di v.a. $\{X_n\}$ che soddisfa le condizioni 1 e 2 ma non la 3. In questo caso non avremo più un'unica matrice di transizione P ma una sequenza di matrici stocastiche $\{P_n\}$, $P_n = [p_n(i,j)]_{i,j\in S}$, una per ciascun passo del processo: per ogni $n \in \mathbb{N}$ e ogni $i, j \in S$

$$\Pr(X_{n+1} = j \mid X_n = i) = p_n(i, j) \ .$$

Infine, un'altra notevole estensione è data dai cosiddetti *processi di Markov*: supponi che il parametro n non sia più discreto ma continuo; allora avremo una famiglia di variabili aleatorie $\{X(t) \mid t \in \mathbb{R}_+\}$, a valori su un insieme S di stati, dotata di una corrispondente famiglia di matrici $\{P(h) = [p_h(i,j)]_{i,j\in S} \mid h \in \mathbb{R}_+\}$. Qui la proprietà di Markov (analoga alla condizione 2) si fa più complessa e viene espressa dalla seguente relazione

$$\Pr(X(t_{n+1}) = j \mid X(t_0) = i_0, \dots, X(t_{n-1}) = i_{n-1}, X(t_n) = i) = \Pr(X(t_{n+1}) = j \mid X(t_n) = i)$$

per ogni sequenza di valori $0 \leq t_0 < t_1 < \cdots < t_n < t_{n+1}$, ogni $i, j \in S$ e ogni n-pla di stati $i_0, i_1, \dots, i_{n-1}$. Mentre la proprietà di omogeneità viene data da

$$\Pr(X(t+h) = j \mid X(t) = i) = p_h(i, j)$$

per ogni $t \geq 0$, $h > 0$ e $i, j \in S$.

2.3 Probabilità di transizione

Sia $\{X_n\}$ una catena di Markov finita e omogenea, con spazio degli stati S, distribuzione iniziale μ e matrice di transizione $P = [p(i,j)]_{i,j\in S}$. Valutiamo ora le probabilità di transizione in un numero di passi $n > 1$, in funzione delle potenze della matrice P.

A tale scopo, per ogni $n \in \mathbb{N}$ denotiamo con $p^{(n)}(i, j)$ i coefficienti della matrice P^n. Ovvero poniamo

$$P^n = [p^{(n)}(i, j)]_{i,j\in S}$$

Nota che $p^{(1)}(i, j) = p(i, j)$, mentre $p^{(0)}(i, j)$ coincide con il coefficiente di Kronecker δ_{ij} e, per il prodotto di matrici, si ha

$$p^{(n+1)}(i, j) = \sum_{\ell\in S} p(i,\ell)p^{(n)}(\ell, j)$$

La proposizione seguente è importante perché stabilisce che $p^{(n)}(i, j)$ coincide con la probabilità di passare dallo stato i allo stato j in n passi in un qualunque momento del processo.

Proposizione 2.1 *Per ogni $k, n \in \mathbb{N}$ e ogni $i, j \in S$, supponendo $Pr_\mu(X_k = i) > 0$ abbiamo*

$$Pr_\mu(X_{k+n} = j \mid X_k = i) = p^{(n)}(i, j)$$

Dimostrazione. Per $n = 0$ la proprietà è ovvia. Se $n = 1$ essa coincide con la condizione 3 della definizione (omogeneità). Ragionando per induzione, supponiamo che la proprietà sia vera per $n \geq 1$. Ricordiamo inoltre che per ogni tripla di eventi A, B, C si ha

$$\Pr(A \cap B \mid C) = \Pr(B \mid C) \cdot \Pr(A \mid B \cap C)$$

Otteniamo così la sequente catena di uguaglianze

$$\begin{aligned}
\Pr_\mu(X_{k+n+1} = j \mid X_k = i) &= \Pr_\mu(X_{k+n+1} = j,\ \exists \ell \in S\ :\ X_{k+1} = \ell \mid X_k = i) \\
&= \sum_{\ell\in S} \Pr_\mu(X_{k+n+1} = j,\ X_{k+1} = \ell \mid X_k = i) \\
&= \sum_{\ell\in S} \Pr_\mu(X_{k+1} = \ell \mid X_k = i) \cdot \Pr_\mu(X_{k+n+1} = j \mid X_{k+1} = \ell,\ X_k = i)
\end{aligned}$$

Nota che $\Pr_\mu(X_{k+1} = \ell \mid X_k = i) > 0$ implica $\Pr_\mu(X_{k+1} = \ell, X_k = i) > 0$. Quindi per la proprietà di Markov si ottiene

$$\begin{aligned}\Pr_\mu(X_{k+n+1} = j \mid X_k = i) &= \sum_{\ell \in S} \Pr_\mu(X_{k+1} = \ell \mid X_k = i) \cdot \Pr_\mu(X_{k+n+1} = j \mid X_{k+1} = \ell) \\ &= \sum_{\ell \in S} p(i,\ell) \cdot \Pr_\mu(X_{k+n+1} = j \mid X_{k+1} = \ell)\end{aligned}$$

che, per ipotesi di induzione, diventa

$$\Pr_\mu(X_{k+n+1} = j \mid X_k = i) = \sum_{\ell \in S} p(i,\ell) \cdot p^{(n)}(\ell, j) = p^{(n+1)}(i,j)$$

□

La proprietà precedente implica inoltre

$$\Pr_\mu(X_{k+n} = j \mid X_k = i) = \Pr_i(X_n = j) = p^{(n)}(i,j)$$

Di conseguenza, assumendo le stesse ipotesi della proposizione precedente otteniamo

$$\Pr_i(X_{k+n} = j) = \sum_{\ell \in S} \Pr_i(X_k = \ell) \cdot \Pr_\ell(X_n = j) \qquad (\forall i, j \in S)$$

equivalente all'identità $P^{k+n} = P^k \cdot P^n$.

Infine, valutiamo la distribuzione della variabile aleatoria X_n per un qualunque $n \in \mathbb{N}$. Possiamo considerare la distribuzione μ di X_0 come un vettore stocastico a indici in S. Analogamente, per ogni $n \in \mathbb{N}$, possiamo rappresentare la distribuzione di X_n mediante il vettore $\mu(n)$ tale che $\mu(n)_j = \Pr_\mu(X_n = j)$ per ogni $j \in S$. Tale vettore verifica la relazione

$$\mu(n)' = \mu' P^n$$

Infatti, per ogni $j \in S$ abbiamo

$$\begin{aligned}\mu(n)_j &= \Pr_\mu(X_n = j) = \sum_{i \in S} \Pr_\mu(X_n = j \mid X_0 = i) \cdot \Pr_\mu(X_0 = i) \\ &= \sum_{i \in S} p^{(n)}(i,j) \cdot \mu_i = (\mu' P^n)_j\end{aligned}$$

Vogliamo ora definire le funzioni generatrici associate ai coefficienti $\{p^{(n)}(i,j)\}_n$. Osserviamo anzitutto che, data una matrice stocastica P, ogni sua potenza P^n, $n \in \mathbb{N}$, è ancora una matrice stocastica e quindi i suoi elementi sono compresi nell'intervallo $[0,1]$. Questo significa che $\lim_{n \to +\infty} (Pz)^n = 0$ per ogni $z \in \mathbb{C}$ tale che $|z| < 1$, e quindi per la proposizione A.1, la matrice $(I - Pz)^{-1} = \sum_{n=0}^{+\infty} P^n z^n$ è ben definita. Possiamo allora dare la seguente definizione.

Definizione 2.1 *La funzione di Green associata a una matrice stocastica P è una matrice di funzioni $G(z)$ data da*

$$G(z) = (I - Pz)^{-1} = \sum_{n=0}^{+\infty} P^n z^n$$

definita per ogni $z \in \mathbb{C}$ tale che $|z| < 1$.

Per ogni $i, j \in S$ possiamo allora considerare la funzione $G_{ij}(z) = ((I - Pz)^{-1})_{ij}$. Per le proprietà presentate nella sezione precedente essa rappresenta la funzione generatrice della sequenza $\{\Pr_i(X_n = j)\}_n$:

$$G_{ij}(z) = \sum_{n=0}^{+\infty} \Pr_i(X_n = j) z^n$$

Poiché $(I - Pz)^{-1} = \frac{\mathrm{adj}(I-Pz)}{\det(I-Pz)}$, ogni funzione $G_{ij}(z)$ è una funzione razionale in z, esistono cioè due polinomi $r(z)$, $s(z)$ a coefficienti in $\mathbb{R}$ (in generale dipendenti da i e j), che soddisfano $\mathrm{gr}(s) < \mathrm{gr}(r)$, tali che

$$G_{ij}(z) = \frac{s(z)}{r(z)}$$

Quindi il problema di determinare la probabilità di transizione in n passi da uno stato all'altro, è ridotto al calcolo del coefficiente n-esimo dello sviluppo in serie di Taylor di una funzione razionale. Più precisamente, per ogni $i, j \in S$ e ogni $n \in \mathbb{N}$, il valore $\Pr_i(X_n = j)$ è il coefficiente n-esimo dello sviluppo in serie di Taylor (con centro in 0) della funzione $\frac{s(z)}{r(z)}$, dove $s(z) = \mathrm{cof}(I - Pz)_{ji}$ e $r(z) = \det(I - Pz)$.

2.4 Tempi di prima entrata

Considera una catena di Markov finita e omogenea $\{X_n\}$, con spazio degli stati S e matrice di transizione $P = [p(i, j)]_{i,j \in S}$. Per ogni $j \in S$ denotiamo con τ_j la variabile aleatoria che rappresenta il minimo $n \geq 1$ tale che $X_n = j$:

$$\tau_j = \min\{n \geq 1 \mid X_n = j\}$$

La variabile τ_j rappresenta quindi il tempo di attesa necessario per entrare in j per la prima volta dopo l'istante 0. Essa assume valori nell'insieme $\mathbb{N}_+ \cup \{+\infty\}$, dove $\mathbb{N}_+ = \{n \in \mathbb{N} \mid n \neq 0\}$. L'evento $\{\tau_j = +\infty\}$ significa che dopo l'istante 0 la catena non entra mai nello stato j, ovvero $X_n \neq j$ per ogni $n \geq 1$.

Definiamo ora i seguenti coefficienti di probabilità per ogni $i, j \in S$:

$$\begin{aligned} f^0(i,j) &= 0 \\ f^{(n)}(i,j) &= \Pr_i(\tau_j = n) \qquad (\forall n \in \mathbb{N}_+) \\ f(i,j) &= \sum_{n \geq 1} f^{(n)}(i,j) = \Pr_i(\tau_j < +\infty) \end{aligned}$$

Nota che la serie precedente è per forza convergente poiché gli eventi $\{\tau_j = n \mid X_0 = i\}$, per $n \geq 1$, sono disgiunti. Chiaramente $f(i, j)$ rappresenta la probabilità di entrare prima o poi in j partendo dallo stato i. Inoltre, è chiaro che

$$\Pr_i(\tau_j = +\infty) = 1 - f(i,j)$$

Nota anche che $f(j, j) = \Pr_j(\exists n \geq 1 : \tau_j = n)$ rappresenta la probabilità di rientrare in j partendo proprio da j, ovvero la probabilità di rientro nel medesimo stato.

Le seguenti osservazioni seguono facilmente dalle definizioni:

1. $f^{(1)}(i, j) = p(i, j)$;
2. $f^{(n)}(i, j) = \Pr_i(X_n = j,\, X_\ell \neq j \;\forall \ell = 1, 2, \ldots, n-1)$;
3. $f(i, j) = \Pr_i(\bigcup_{n \geq 1}\{X_n = j\})$;
4. $\sup_{n \geq 1}\{p^{(n)}(i, j)\} \leq f(i, j) \leq \sum_{n \geq 1} p^{(n)}(i, j)$.

I coefficienti $f^{(n)}(i,j)$ sono inoltre legati alle probabilità di transizione in più passi. La seguente proprietà è abbastanza intuitiva: essa afferma che per andare da i a j in n passi occorre entrare per la prima volta in j alla k-esima mossa, per qualche $k \in \{1,\dots,n\}$ e quindi rientrare in j nei successivi $n-k$ passi.

Proposizione 2.2 *Per ogni $n \geq 1$ e ogni $i,j \in S$*

$$p^{(n)}(i,j) = \sum_{k=1}^{n} f^{(k)}(i,j)p^{(n-k)}(j,j) \tag{2.1}$$

Dimostrazione. Condizionando sull'evento $\{\tau_j = k\}$, $k = 1,\dots,n$, abbiamo

$$p^{(n)}(i,j) = \mathrm{Pr}_i(X_n = j) = \sum_{k=1}^{n} \mathrm{Pr}_i(X_n = j \mid \tau_j = k)\mathrm{Pr}_i(\tau_j = k) \tag{2.2}$$

Se $\mathrm{Pr}_i(\tau_j = k) \neq 0$, per la proprietà di Markov e la proposizione 2.1, vale $\mathrm{Pr}_i(X_n = j \mid \tau_j = k) = \mathrm{Pr}_i(X_n = j \mid X_k = j) = p^{(n-k)}(j,j)$. La proprietà segue quindi dalla equazione (2.2) sostituendo opportunamente i valori nella sommatoria. □

La proposizione precedente stabilisce una convoluzione tra sequenze che quindi può essere trasformata in un prodotto di funzioni generatrici. A tale scopo definiamo, per ogni $i,j \in S$, la funzione generatrice $F_{ij}(z)$ ponendo

$$F_{ij}(z) = \sum_{n=0}^{+\infty} f^{(n)}(i,j)z^n$$

Poiché $F_{ij}(1) = f(i,j)$, la serie converge per $z = 1$ e quindi il suo raggio di convergenza è maggiore o uguale a 1.

Ora, osserviamo che se $i \neq j$ allora $p^{(0)}(i,j) = 0 = f^{(0)}(i,j)p^{(0)}(j,j)$. Di conseguenza, per la proposizione 2.2, la sequenza $\{p^{(n)}(i,j)\}_{n\geq 0}$ è la convoluzione di $\{f^{(n)}(i,j)\}_{n\geq 0}$ e di $\{p^{(n)}(j,j)\}_{n\geq 0}$. Questo implica

$$G_{ij}(z) = F_{ij}(z)G_{jj}(z)$$

Se invece $i = j$ allora $p^{(0)}(i,j) = 1$ e per lo stesso ragionamento otteniamo

$$G_{jj}(z) = 1 + F_{jj}(z)G_{jj}(z)$$

Abbiamo così dimostrato il seguente corollario.

Corollario 2.3 *Per ogni $i,j \in S$, se $i \neq j$ allora*

$$G_{ij}(z) = F_{ij}(z)G_{jj}(z) \qquad (|z| < 1)$$

altrimenti abbiamo

$$G_{jj}(z) = \frac{1}{1 - F_{jj}(z)} \qquad (|z| < 1)$$

L'ultima proprietà che presentiamo in questa sezione permette di esprimere attraverso i coefficienti $f(i,j)$ la probabilità di entrare in un dato stato un certo numero di volte.

Proposizione 2.4 *Per ogni $i,j \in S$ e ogni intero $k \geq 1$ vale la seguente relazione*

$$Pr_i(X_n = j \text{ per almeno } k \text{ valori } n > 0) = f(i,j)\left(f(j,j)\right)^{k-1}$$

Dimostrazione. Ragioniamo per induzione su $k \geq 1$. Se $k = 1$ la proprietà è ovvia. Supponiamo l'equazione vera per $k > 1$ e dimostriamola per $k+1$. Condizionando sul valore di τ_j abbiamo

$$\begin{aligned}
&\Pr_i(X_n = j \text{ per almeno } k+1 \text{ valori } n > 0) \\
&\quad = \sum_{t\geq 1} \Pr_i(\tau_j = t,\ X_n = j \text{ per almeno } k \text{ valori } n > t) \\
&\quad = \sum_{t\geq 1} \Pr_i(X_n = j \text{ per almeno } k \text{ valori } n > t \mid \tau_j = t) \Pr_i(\tau_j = t) \\
&\quad = \sum_{t\geq 1} \Pr_j(X_n = j \text{ per almeno } k \text{ valori } n > 0) \Pr_i(\tau_j = t) \\
&\quad = \Pr_j(X_n = j \text{ per almeno } k \text{ valori } n > 0) \sum_{t\geq 1} \Pr_i(\tau_j = t) \\
&\quad = (f(j,j))^k f(i,j)
\end{aligned}$$

dove la proprietà di Markov e l'omogeneità sono utilizzate nella terza uguaglianza, mentre l'ipotesi di induzione nell'ultima. □

2.5 Stati ricorrenti

Sia di nuovo $\{X_n\}$ una catena di Markov finita e omogenea, con uno spazio degli stati S e matrice di transizione $P = [p(i,j)]_{i,j\in S}$. Diciamo che uno stato $i \in S$ è *ricorrente* se

$$\Pr_i(X_n = i \text{ per infiniti } n > 0) = 1$$

Dimostriamo innazitutto che la definizione è ben posta, cioé che l'evento

$$\{X_n = i \text{ per infiniti } n > 0\}$$

appartiene alla σ-algebra $\mathcal{B}$. Infatti, per ogni intero $k \geq 1$, l'evento

$$\alpha_k = \{X_n = i \text{ per almeno } k \text{ valori } n > 0\}$$

appartiene a $\mathcal{B}$ (essendo ottenibile mediante unione di cilindri). Inoltre, si vede subito che

$$\alpha_1 \supseteq \alpha_2 \supseteq \cdots \supseteq \alpha_k \supseteq \alpha_{k+1} \supseteq \cdots$$

e che

$$\{X_n = i \text{ per infiniti } n > 0\} = \bigcap_{k\geq 1} \alpha_k$$

Di conseguenza anche $\{X_n = i \text{ per infiniti } n > 0\}$ appartiene a $\mathcal{B}$ e inoltre

$$\Pr_i(X_n = i \text{ per infiniti } n > 0) = \lim_{k\to+\infty} \Pr_i(\alpha_k) \tag{2.3}$$

Diremo inoltre che uno stato $i \in S$ è *transiente* se non è ricorrente.

Proposizione 2.5 *Per ogni stato $i \in S$ le seguenti proprietà sono equivalenti:*

1. *i è ricorrente;*
2. *$f(i,i) = 1$;*
3. *$\sum_{n\geq 0} p^{(n)}(i,i) = +\infty$.*

Dimostrazione. Per provare che la prima proprietà equivale alla seconda osserva che, per la relazione (2.3) e la proposizione 2.4, vale

$$\Pr_i(X_n = i \text{ per infiniti } n > 0) = \lim_{k\to+\infty} (f(i,i))^k \tag{2.4}$$

e tale limite è 1 se $f(i,i) = 1$, mentre è 0 se $f(i,i) < 1$.

Per dimostrare l'equivalenza tra la seconda e la terza proprietà ricordiamo innazitutto il teorema di Abel. Questo afferma che data una serie di potenze $A(z) = \sum_{n\geq 0} a_n z^n$, con raggio di convergenza $R \geq 1$ e coefficienti $a_n \geq 0$ (per ogni $n \in \mathbb{N}$), vale

$$\lim_{z\to 1^-,\, z\in[0,1]} A(z) = A(1)$$

dove il limite è inteso per z che si muove sull'intervallo reale $[0,1]$ e approssima 1 per difetto, mentre $A(1) = \sum_{n\geq 0} a_n$ può anche assumere il valore $+\infty$.

Ora, osserva che per ogni stato i, $f(i,i) = F_{ii}(1)$ è sempre minore o uguale a 1. Se $f(i,i) = 1$, per il teorema di Abel abbiamo

$$1 = F_{ii}(1) = \lim_{z\to 1^-,\, z\in[0,1]} F_{ii}(z)$$

e quindi per il corollario 2.3,

$$G_{ii}(1) = \lim_{z\to 1^-,\, z\in[0,1]} \frac{1}{1 - F_{ii}(z)} = +\infty$$

ovvero $G_{ii}(1) = \sum_{n\geq 0} p^{(n)}(i,i) = +\infty$.

Viceversa, se $f(i,i) < 1$ per lo stesso teorema abbiamo $\lim_{z\to 1^-,\, z\in[0,1]} F_{ii}(z) < 1$ e quindi

$$G_{ii}(1) = \lim_{z\to 1^-,\, z\in[0,1]} \frac{1}{1 - F_{ii}(z)} < +\infty$$

che implica la convergenza della serie $G_{ii}(1) = \sum_{n\geq 0} p^{(n)}(i,i)$. □

Osservazione 2 Nota che per la (2.4) l'evento $\{X_n = i \text{ per infiniti } n > 0\}$ obbedisce a una legge 0-1, nel senso che la sua probabilità è 0 oppure 1. Inoltre, un enunciato analogo vale per gli stati transienti: uno stato i è transiente se e solo se $f(i,i) < 1$ ovvero, se e solo se la serie $\sum_{n\geq 0} p^{(n)}(i,i)$ converge.

Osserviamo che quest'ultima proprietà può essere ulteriormente estesa. Infatti, per ogni $i, j \in S$, se j è transiente allora

$$\sum_{n\geq 0} p^{(n)}(i,j) < +\infty \tag{2.5}$$

La prova è ancora basata sul teorema di Abel ed è analoga a quella della proposizione precedente (unica differenza: si tiene conto dell'equazione $G_{ij}(z) = F_{ij}(z)G_{jj}(z)$). Tuttavia non vale il viceversa, ovvero la convergenza della serie non implica che lo stato j sia transiente.

Concludiamo questa sezione mostrando un legame interessante tra la proprietà di ricorrenza ora introdotta e le classi di irriducibilità del grafo associato alla catena di Markov considerata.

Proposizione 2.6 *La ricorrenza è una proprietà delle classi. Ovvero, se $i \in S$ è uno stato ricorrente e C è la sua classe allora ogni stato j in C è ricorrente.*

Dimostrazione. Data una classe C, siano $i, j \in C$ due stati distinti. Allora esistono $a, b > 0$, $a = p^{(r)}(i, j)$ e $b = p^{(s)}(j, i)$ per qualche $r, s \in \mathbb{N}$, tali che per ogni $n \in \mathbb{N}$

$$p^{(n+r+s)}(j, j) \geq b\, a\, p^{(n)}(i, i) \tag{2.6}$$

Per la proposizione 2.5, se i è ricorrente la serie $\sum_{n\geq 0} p^{(n)}(i, i)$ è divergente e per la (2.6) anche $\sum_{n\geq 0} p^{(n)}(j, j)$ diverge, per cui j risulta ricorrente. Nota che per la simmetria di i e j, se i fosse transiente anche j lo sarebbe. □

In parole povere questo significa che in ogni classe tutti gli stati sono ricorrenti, oppure sono tutti transienti.

2.6 Equivalenza tra stati ricorrenti e stati essenziali

In questa sezione proviamo (sempre nell'ipotesi di catene di Markov finite e omogenee) che gli stati ricorrenti coincidono con quelli essenziali. Ricordiamo che in una catena di Markov uno stato i è essenziale se per ogni stato j, $i \to j$ implica $j \to i$ (ovvero se $p^{(n)}(i, j) > 0$ implica $p^{(k)}(j, i) > 0$ per qualche $k \in \mathbb{N}$). Quindi la nozione di ricorrenza, definita in modo probabilistico, coincide con una proprietà dei nodi del grafo di connessione della catena.

Consideriamo dunque una catena di Markov $\{X_n\}$ su un insieme finito di stati S con matrice di transizione $P = [p(i, j)]_{i,j\in S}$.

Proposizione 2.7 *Se uno stato i è ricorrente e $i \to j$ allora anche j è ricorrente e inoltre $f(i, j) = 1 = f(j, i)$.*

Dimostrazione. Mostriamo anzitutto che $f(j, i) = 1$. Poiché i è ricorrente, per ogni $n \in \mathbb{N}$ abbiamo

$$\begin{aligned} 1 &= \Pr_i(X_m = i \text{ per infiniti } m > 0) \leq \Pr_i(X_m = i \text{ per qualche } m > n) = \\ &= \sum_{r\in S} \Pr_i(X_m = i \text{ per qualche } m > n,\ X_n = r) \\ &= \sum_{r\in S} p^{(n)}(i, r) \Pr_i(X_m = i \text{ per qualche } m > n \mid X_n = r) \\ &= \sum_{r\in S} p^{(n)}(i, r)\, f(r, i) \end{aligned}$$

Scegliamo ora n tale che $p^{(n)}(i, j) > 0$. Se fosse $f(j, i) < 1$ allora avremmo

$$1 \leq \sum_{r\in S} p^{(n)}(i, r)\, f(r, i) < \sum_{r\in S} p^{(n)}(i, r) = 1$$

che è assurdo. Quindi $f(j, i) = 1$ e questo implica che $j \to i$. Di conseguenza i e j appartengono alla stessa classe e per la proposizione 2.6 anche j è ricorrente. Scambiando infine i e j nel ragionamento precedente otteniamo $f(i, j) = 1$. □

Corollario 2.8 *Ogni stato ricorrente è essenziale.*

Dimostrazione. Infatti nella prova della proposizione 2.7 abbiamo visto che se i è uno stato ricorrente e $i \to j$ allora $j \to i$. Questo significa che i è essenziale. □

Proposizione 2.9 *In una catena di Markov finita e omogenea ogni stato essenziale è ricorrente.*

Dimostrazione. Sia i uno stato essenziale e sia C la sua classe. Ragioniamo per assurdo e supponiamo che i sia transiente. Allora tutti gli stati $j \in C$ sono transienti. Quindi, per la (2.5), otteniamo $\sum_{n\geq 0} p^{(n)}(i,j) < +\infty$ e di conseguenza $\lim_{n\to+\infty} p^{(n)}(i,j) = 0$ per ogni $j \in C$.

Inoltre, poiché C è essenziale, per ogni $n \in \mathbb{N}$ abbiamo

$$\sum_{j\in C} p^{(n)}(i,j) = 1$$

Così, possiamo scrivere

$$1 = \lim_{n\to+\infty} \sum_{j\in C} p^{(n)}(i,j) = \sum_{j\in C} \lim_{n\to+\infty} p^{(n)}(i,j) \tag{2.7}$$

La seconda uguaglianza è giustificata dal fatto che C è finita e che, per ogni $j \in C$, esiste finito il limite $\lim_n p^{(n)}(i,j)$. Tuttavia tali limiti sono tutti nulli e sostituendo i loro valori in (2.7) otteniamo un assurdo (cioè $1 = 0$). Ne segue che i deve essere ricorrente. □

Osservazione 3 La dimostrazione precedente è basata sullo scambio tra limite e somma nella (2.7), del tipo

$$\lim_{n\to+\infty} \sum_{j\in C} a_n(j) = \sum_{j\in C} \lim_{n\to+\infty} a_n(j) \qquad (\{a_n(j)\}_n \subseteq \mathbb{R}_+ \text{ per ogni } j \in C)$$

Questa uguaglianza è sempre vera se C è finito e i limiti del termine destro esistono e sono finiti. Se invece C è un insieme infinito la proprietà non è sempre vera. Per esempio:

$$\lim_{n\to+\infty} \sum_{j=0}^{+\infty} \frac{n^j}{j!e^n} = \lim_{n\to+\infty} e^{-n} \sum_{j=0}^{+\infty} \frac{n^j}{j!} = 1$$

mentre

$$\sum_{j=0}^{+\infty} \lim_{n\to+\infty} \frac{n^j}{j!e^n} = 0$$

Il noto teorema di Lebesgue di convergenza dominata fornisce condizioni che permettono lo scambio tra limite e sommatoria anche con infiniti elementi (vedi per esempio [5]).

Corollario 2.10 *Per ogni $i, j \in S$, se j è transiente allora*

$$p^{(n)}(i,j) = O(\varepsilon^n)$$

per qualche $0 < \varepsilon < 1$.

Dimostrazione. Ricordiamo anzitutto che per le proposizioni precedenti uno stato è ricorrente se e solo se è essenziale. Quindi, se i è ricorrente la proprietà è ovvia perché $p^{(n)}(i,j) = 0$ per ogni n. Sia allora $T \subseteq S$ l'insieme degli stati transienti e supponi $i \in T$. Sappiamo che per ogni $\ell \in T$ esiste uno stato ricorrente v tale che $\ell \to v$. Di conseguenza esistono un $t \in \mathbb{N}$ e un numero reale α, $0 < \alpha < 1$, tali che per ogni $\ell \in T$

$$\sum_{u\in T} p^{(t)}(\ell,u) \leq \alpha$$

Ragionando per induzione si può provare che per ogni intero $s > 0$ e ogni $\ell \in T$

$$\sum_{u \in T} p^{(st)}(\ell, u) = \sum_{u,v \in T} p^{(t)}(\ell, v) p^{((s-1)t)}(v, u) \leq \sum_{v \in T} p^{(t)}(\ell, v) \alpha^{s-1} \leq \alpha^s$$

In particolare questo implica

$$\sum_{u \in T} p^{(st)}(i, u) \leq \alpha^s$$

Se ora consideriamo un $n \in \mathbb{N}$ qualsiasi e poniamo $n = st + r$, dove r è il resto della divisione di n per t, abbiamo

$$p^{(n)}(i, j) = \sum_{u \in T} p^{(st)}(i, u)\; p^{(r)}(u, j) \leq \alpha^s = \alpha^{\frac{n-r}{t}} = \mathrm{O}(\alpha^{1/t})^n$$

Poiché $0 < \alpha^{1/t} < 1$ la proprietà è dimostrata. □

Nella prossima proposizione si prova che quasi certamente ogni catena di Markov finita entra definitivamente in una classe essenziale.

Proposizione 2.11 *Con probabilità* 1 *ogni catena di Markov (finita e omogenea) esce definitivamente dall'insieme T degli stati transienti. Ovvero, per ogni stato i,*

$$Pr_i(\exists k \in \mathbb{N} \,:\, X_n \notin T \textit{ per ogni } n > k) = 1$$

Dimostrazione. Poiché T è finito, se $X_n \in T$ per ogni $n \in \mathbb{N}$ allora esiste $j \in T$ tale che $X_n = j$ per infiniti n. Quindi, per ogni $i \in S$, abbiamo

$$\begin{aligned}
\mathrm{Pr}_i(X_n \in T \text{ per ogni } n \in \mathbb{N}) &\leq \mathrm{Pr}_i(\exists j \in T \,:\, X_n = j \text{ per infiniti } n) \\
&\leq \sum_{j \in T} \mathrm{Pr}_i(X_n = j \text{ per infiniti } n) \\
&= \sum_{j \in T} f(i, j) \mathrm{Pr}_j(X_n = j \text{ per infiniti } n)
\end{aligned}$$

Per l'osservazione 2, l'evento $(X_n = j \text{ per infiniti } n)$ obbedisce a una legge $0 - 1$ e inoltre, se j è transiente, la sua probabilità non può essere 1. Quindi $\mathrm{Pr}_j(X_n = j \text{ per infiniti } n) = 0$ per ogni $j \in T$ e di conseguenza dall'ultima disuguaglianza si ricava

$$\mathrm{Pr}_i(X_n \in T \text{ per ogni } n \in \mathbb{N}) = 0$$

Poiché T coincide con l'insieme degli stati non essenziali, una volta uscita da T la catena non vi rientra più. Ne segue che

$$\begin{aligned}
\mathrm{Pr}_i(\exists k \in \mathbb{N} \,:\, X_n \notin T \text{ per ogni } n > k) &= \mathrm{Pr}_i(\exists m \in \mathbb{N} \,:\, X_m \notin T) \\
&= 1 - \mathrm{Pr}_i(X_n \in T \text{ per ogni } n \in \mathbb{N}) \;=\; 1
\end{aligned}$$

□

2.7 Tempi medi di rientro

In questa sezione studiamo i tempi medi di rientro in un dato stato. Come al solito, sia $\{X_n\}$ una catena di Markov finita e omogenea, con insieme degli stati S e matrice di transizione $P = [p(i,j)]_{i,j\in S}$. Ricordando la definizione data nella sezione 2.4, per ogni $i \in S$ sia τ_i il tempo di prima entrata nello stato i successivo all'istante 0. Vogliamo valutare il valor medio di τ_i supponendo che $X_0 = i$, ovvero il valore

$$E_i(\tau_i) = \sum_{n\in\mathbb{N}_+\cup\{+\infty\}} n \Pr_i(\tau_i = n)$$

Consideriamo ovviamente il caso in cui lo stato i sia ricorrente; se infatti i fosse transiente, avremmo $E_i(\tau_i) = +\infty$ poiché $\Pr_i(\tau_i = \infty) > 0$.

Se i è ricorrente allora

$$E_i(\tau_i) = \sum_{n\geq 1} n \Pr_i(\tau_i = n) = \sum_{n\geq 1} n\, f^{(n)}(i,i)$$

e l'obiettivo principale di questa sezione è quello di provare che in questo caso $E_i(\tau_i)$ è finito.

Osservazione 4 A questo proposito è bene ricordare che esistono naturali variabili aleatorie a valori in $\mathbb{N}$ con valor medio infinito. Per esempio, considera una variabile aleatoria X a valori in $\mathbb{N}$ tale che

$$\Pr(X = n) = \begin{cases} 2^{-i} & \text{se } n = 2^i \text{ per qualche intero } i \geq 1 \\ 0 & \text{altrimenti} \end{cases}$$

e osserva che $E(X) = \sum_{i\geq 1} 2^i \cdot 2^{-i} = \sum_{i\geq 1} 1 = +\infty$.

Un esempio significativo di questo tipo è dato dal paradosso di San Pietroburgo, illustrato in [3].

Sappiamo già dalla sezione 2.4 che la serie di potenze $F_{ii}(z) = \sum_{n\geq 1} f^{(n)}(i,i) z^n$ possiede raggio di convergenza $r \geq 1$. Consideriamo allora la sua derivata $F'_{ii}(z)$:

$$F'_{ii}(z) = \sum_{n\geq 1} n f^{(n)}(i,i) z^{n-1}$$

È facile verificare che il raggio di convergenza di $F'_{ii}(z)$ coincide con quello di $F_{ii}(z)$ ed è quindi maggiore o uguale a 1. Di conseguenza, per il teorema di Abel abbiamo

$$E_i(\tau_i) = F'_{ii}(1) = \lim_{z\to 1^-,\, z\in[0,1]} F'_{ii}(z)$$

Questo significa che per provare $E_i(\tau_i) < +\infty$ è sufficiente dimostrare che $f^{(n)}(i,i) = \mathrm{O}(\varepsilon^n)$ per qualche $0 < \varepsilon < 1$; in questo caso infatti il raggio di convergenza di $F'_{ii}(z)$ risulta maggiore di 1 e quindi $F'_{ii}(1) = E_i(\tau_i) \in \mathbb{R}_+$.

Proposizione 2.12 *Per ogni $i \in S$ ricorrente abbiamo $E_i(\tau_i) < +\infty$.*

Dimostrazione. Per l'osservazione precedente ci limitiamo a provare che $f^{(n)}(i,i) = \mathrm{O}(\varepsilon^n)$ per qualche $0 < \varepsilon < 1$. Inoltre, essendo i essenziale, possiamo restringere l'insieme degli stati S alla classe C di i (infatti nessuno stato fuori da C è raggiungibile da i). Definiamo allora una nuova catena di Markov sull'insieme degli stati C rendendo i assorbente. Tale catena possiede matrice di transizione $\tilde{P} = [\tilde{p}(k,j)]_{k,j\in C}$ tale che

$$\begin{aligned} \tilde{p}(k,j) &= p(k,j) \qquad \text{per ogni } k,j \in C \text{ con } k \neq i \\ \tilde{p}(i,j) &= \begin{cases} 1 & \text{se } j = i \\ 0 & \text{altrimenti} \end{cases} \end{aligned}$$

Nella nuova catena tutti gli stati $j \neq i$ sono transienti e quindi, per il corollario 2.10,

$$\sum_{j \neq i} \tilde{p}^{(n)}(k,j) = O(\varepsilon^n)$$

per qualche $0 < \varepsilon < 1$ (e per tutti i $k \in C$). Ne segue allora che per ogni intero $n \geq 2$

$$\begin{aligned} f^{(n)}(i,i) &= \sum_{k,j \in C,\, k,j \neq i} p(i,k)\, \tilde{p}^{(n-2)}(k,j)\, p(j,i) \\ &= O(\varepsilon^n) \sum_{k,j \in C,\, k,j \neq i} p(i,k)\, p(j,i) \\ &= O(\varepsilon^n) \end{aligned}$$

□

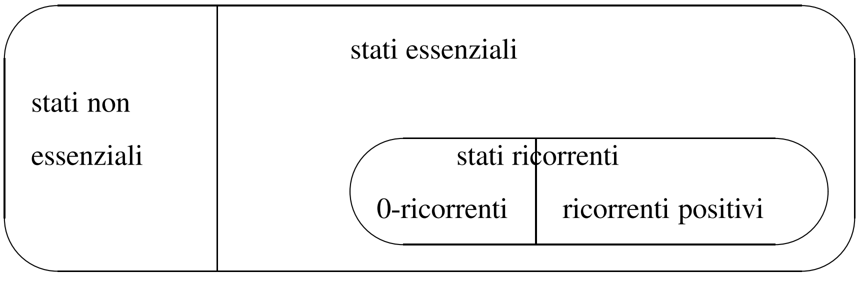
2.8 Catene di Markov infinite

A questo punto una domanda che sorge spontanea è la seguente: cosa cambia se l'insieme degli stati S è infinito (numerabile)? Osserviamo infatti che una catena di Markov può essere definita come nella sezione 2.2 anche nel caso in cui l'insieme degli stati S sia infinito e numerabile. Le probabilità di transizioni in più passi si calcolano nello stesso modo, le nozioni di stato ricorrente, transiente, essenziale, nonché i coefficienti $f(i,j)$ e $f^{(n)}(i,j)$ si possono definire in maniera analoga. Tuttavia in questo caso molte delle dimostrazioni presentate nelle sezioni precedenti non sono più vere. Le principali differenze sono le seguenti:

1. Ogni stato ricorrente è anche uno stato essenziale; tuttavia possono esistere stati essenziali che non sono ricorrenti. Quindi in alcuni casi la famiglia degli stati ricorrenti è propriamente inclusa in quella degli stati essenziali.

2. Non è detto che uno stato ricorrente abbia tempo medio di rientro finito: possono esistere stati ricorrenti i tali che $E_i(\tau_i) = +\infty$ (i cosiddetti stati 0-ricorrenti).

3. Non è detto che la catena termini in una classe essenziale. Ovvero può capitare che con probabilità maggiore di 0 la catena rimanga per sempre in stati non essenziali.



In questo contesto, uno stato ricorrente i si dice *ricorrente positivo* se $E_i(\tau_i) < +\infty$, mentre i è chiamato 0-*ricorrente* se $E_i(\tau_i) = +\infty$.

Riassumiamo qui di seguito le principali proprietà:

a) Uno stato i è ricorrente se e solo se $G_{ii}(1) = \sum_{n \geq 0} p^{(n)}(i,i) = +\infty$, ovvero se e solo se $F_{ii}(1) = 1$.

b) Ogni stato ricorrente è essenziale.

c) La ricorrenza è una proprietà delle classi, così come la 0-ricorrenza e la ricorrenza positiva.

d) Ogni classe essenziale finita è anche una classe ricorrente positiva. Come abbiamo segnalato sopra la proprietà inversa non è vera. In particolare esistono classi essenziali che non sono ricorrenti.

Esempio 2.1 *Passeggiata infinita dell'ubriaco* [11]. Consideriamo una catena di Markov sullo spazio degli stati $\mathbb{Z}$, con probabilità di transizione definite mediante due valori $p,q > 0$, $p+q = 1$, ponendo per ogni $k \in \mathbb{Z}$

$$p(k,k+1) = p, \quad p(k,k-1) = q \ \text{ e } \ p(k,i) = 0 \ \text{ per ogni intero } i \neq k+1, k-1.$$

È chiaro che $\mathbb{Z}$ forma un'unica classe essenziale. Inoltre, si può provare che

$$\begin{aligned} F_{00}(z) &= \sum_{n\geq 0} f^{(n)}(0,0) z^n = 1 - \sqrt{1-4qpz^2} \\ F_{00}(1) &= f(0,0) = 1 - |p-q| = \begin{cases} 1 & \text{se } p = q \\ 2\min\{p,q\} & \text{se } p \neq q \end{cases} \end{aligned}$$

Quindi lo stato 0 è ricorrente se $p = q$, mentre è transiente se $p \neq q$. Di conseguenza nel primo caso tutti gli stati sono ricorrenti, nel secondo sono tutti transienti.

Inoltre, se $p = q$ si verifica che tutti gli stati sono 0-ricorrenti. Infatti in questo caso abbiamo

$$F_{00}(z) = 1 - \sqrt{1-z^2}, \qquad F'_{00}(z) = \frac{z}{\sqrt{1-z^2}}$$

e, per il teorema di Abel,

$$E_0(\tau_0) = \lim_{z\to 1^-} F'_{00}(z) = +\infty$$

da cui si ricava che lo stato 0 è 0-ricorrente.

Modificando l'esempio precedente si può inoltre ottenere una catena di Markov infinita nella quale non è garantita l'entrata in una classe essenziale.

Esempio 2.2 *Passeggiata infinita con barriera assorbente* [11]. Sia $\mathbb{N}$ lo spazio degli stati e, dati $p,q > 0$, $p+q = 1$, definiamo le seguenti probabilità di transizione:

$$p(0,0) = 1, \quad p(k,k+1) = p, \quad p(k,k-1) = q \quad \text{per ogni intero } k > 0$$

mentre $p(i,j) = 0$ per gli altri valori i,j in $\mathbb{N}$. Nella corrispondente catena di Markov, lo stato 0 è assorbente (quindi ricorrente ed essenziale) mentre la famiglia degli interi positivi forma una classe non essenziale infinita.

Inoltre si può dimostrare che

$$F_{10}(z) = \frac{1-\sqrt{1-4pqz^2}}{2pz}, \ \text{da cui} \ \ f(1,0) = \min\{1, q/p\}.$$

Nel caso $p > q$, se lo stato iniziale è 1 la catena finirà nell'unico stato essenziale con probabilità $f(1,0) < 1$. Con probabilità non nulla la catena (partendo dallo stato 1) può rimanere per sempre nella classe non essenziale degli stati $i > 0$.

Esercizi

1) Data una catena di Markov finita, sia C una classe ricorrente. Dimostrare che per ogni coppia di stati $i, j \in C$, $f(i, j) = 1$.

2) Nelle stesse ipotesi dell'esercizio precedente, provare che $E_i(\tau_j) < +\infty$ per ogni $i, j \in C$.

Capitolo 3

Catene di Markov ergodiche

In questo capitolo studiamo le proprietà ergodiche delle catene di Markov finite. Ricordiamo innazitutto la definizione principale.

Una catena di Markov finita $\{X_n\}$ su un insieme di stati S si dice *ergodica* se esiste un vettore stocastico $\pi^* = (\pi_i^*)_{i\in S}$ tale che, per ogni $i, j \in S$,

$$\lim_{n\to+\infty} \mathrm{Pr}_i(X_n = j) \ = \ \pi_j^*$$

ovvero

$$\lim_{n\to+\infty} p^{(n)}(i,j) \ = \ \pi_j^*$$

dove $P = [p(i,j)]_{i,j\in S}$ è la matrice di transizione della catena.

In altre parole, una catena di Markov $\{X_n\}$ è ergodica se al crescere di n esiste la distribuzione limite della variabile X_n ed è indipendente dalla distribuzione iniziale μ. Questa proprietà è essenziale in molte applicazioni e rappresenta uno degli aspetti più studiati nella teoria delle catene di Markov.

L'esempio fondamentale di catena ergodica è fornito dalle catene di Markov primitive. Sappiamo infatti, per la proposizione 1.12, che se P è una matrice stocastica primitiva, allora

$$\lim_{n\to+\infty} P^n = eu' \ ,$$

dove e è il vettore i cui elementi sono uguali a 1 e u è il vettore stocastico che è anche autovettore sinistro di P corrispodente all'autovalore 1. Di conseguenza $\lim_{n\to+\infty} p^{(n)}(i,j) = u_j$ per ogni indice j. Quindi, se $\{X_n\}$ è una catena di Markov con matrice di transizione P, l'autovettore sinistro u rappresenta la distribuzione limite di $\{X_n\}$ ed è indipendente dalla distribuzione iniziale di $\{X_n\}$.

In questo capitolo invece studiamo le proprietà ergodiche delle catene di Markov senza l'ausilio del teorema di Perron-Frobenius e quindi di fatto, senza usare la proposizione 1.12. Svilupperemo un'analisi diretta della sequenze $\{p^{(n)}(i,j)\}_n$ che permetterà di mettere in luce il significato probabilistico della distribuzione limite π^* quando questa esiste. In questo caso infatti avremo che per ogni stato i, $\pi_i^* = 1/E_i(\tau_i)$, dove τ_i è il tempo di prima entrata nello stato i studiato nella sezione 2.4.

3.1 Distribuzioni stazionarie

Data una catena di Markov finita $\{X_n\}$ sull'insieme degli stati $S = \{1,2,\ldots,k\}$, con matrice di transizione P, chiamiamo *distribuzione stazionaria* un vettore $\pi \in \mathbb{R}^k$ che soddisfa le seguenti proprietà:

1. π è un vettore stocastico, ovvero $\pi_i \geq 0$ per ogni $i = 1, 2, \ldots, k$ e inoltre $\sum_{i=1}^{k} \pi_i = 1$.
2. π è autovettore sinistro di P corrispondente a 1, ovvero $\pi' P = \pi'$.

Chiaramente, un tale vettore π verifica l'equazione $\pi' P^n = \pi'$ per ogni intero $n > 0$ e quindi $\Pr_\pi(X_n = i) = \pi_i$ per ogni stato i. Se dunque π coincide con la distribuzione iniziale della catena, la probabilità di trovarsi nei vari stati rimane sempre la stessa durante l'evoluzione del processo.

Proviamo ora a costruire una distribuzione stazionaria per una generica catena di Markov $\{X_n\}$ con famiglia degli stati $S = \{1, 2, \ldots, k\}$ e matrice di transizione P. Poiché ogni catena finita ammette sempre una classe ricorrente, senza perdita di generalità possiamo supporre che lo stato 1 sia ricorrente. Allora, per ogni $i \in S$, definiamo

$$\rho_i = \sum_{n=0}^{+\infty} \Pr_1(X_n = i,\ \tau_1 > n)$$

Tali coefficienti godono delle seguenti proprietà:

1. Ogni ρ_i è finito.

   Infatti abbiamo $\rho_i \leq \sum_{n=0}^{+\infty} \Pr_1(\tau_1 > n) = E_1(\tau_1)$ ed essendo lo stato 1 ricorrente, $E_1(\tau_1) < +\infty$ per la proposizione 2.12.

2. $\rho_1 = 1$.

   Infatti $\Pr_1(X_0 = 1,\ \tau_1 > 0) = 1$, mentre $\Pr_1(X_n = i,\ \tau_1 > n) = 0$ per ogni intero $n > 0$.

3. Ogni ρ_i rappresenta il numero medio di entrate nello stato i nell'intervallo di tempo $\{0, 1, \ldots, \tau_1 - 1\}$, supponendo $X_0 = 1$.

   Per verificare anche questa proprietà, considera la variabile aleatoria $\nu_i = \sum_{n=0}^{+\infty} b_i(n)$, dove

   $$b_i(n) = \begin{cases} 1 & \text{se } X_n = i \text{ e } \tau_1 > n \\ 0 & \text{altrimenti} \end{cases}$$

   Allora, il valor medio di ν_i nell'ipotesi $X_0 = 1$ è dato da

   $$E_1(\nu_i) = \sum_{n=0}^{+\infty} E_1(b_i(n)) = \sum_{n=0}^{+\infty} \Pr_1(X_n = i,\ \tau_1 > n) = \rho_i$$

Definiamo ora il vettore π dato da

$$\pi = \left(\frac{\rho_1}{E_1(\tau_1)}, \frac{\rho_2}{E_1(\tau_1)}, \ldots, \frac{\rho_k}{E_1(\tau_1)} \right) \tag{3.1}$$

Proposizione 3.1 *Il vettore π è stocastico.*

Dimostrazione. Chiaramente $\pi_i \geq 0$ per ogni i. Inoltre abbiamo

$$\begin{aligned} \sum_{i=1}^{k} \rho_i &= \sum_{i=1}^{k} \sum_{n=0}^{+\infty} \Pr_1(X_n = i,\ \tau_1 > n) = \sum_{n=0}^{+\infty} \sum_{i=1}^{k} \Pr_1(X_n = i,\ \tau_1 > n) \\ &= \sum_{n=0}^{+\infty} \Pr_1(\tau_1 > n) = E_1(\tau_1) \end{aligned}$$

□

Il prossimo enunciato stabilisce la proprietà principale di questa sezione.

Proposizione 3.2 *Sia* $\{X_n\}$ *una catena di Markov sugli stati* $\{1,2,\ldots,k\}$ *con matrice di transizione* $P = [p(i,j)]$ *e supponiamo che lo stato* 1 *sia ricorrente. Allora il vettore* π *definito in (3.1) è una distribuzione stazionaria per* $\{X_n\}$.

Dimostrazione. Basta provare che il vettore $\rho = (\rho_1, \rho_2 \ldots, \rho_k)$ è autovettore sinistro di P corrispondente all'autovalore 1. Consideriamo prima il caso $j \neq 1$ e proviamo che $\rho_j = (\rho' P)_j$. A tale scopo si può manipolare la stessa definizione di ρ_j condizionando sullo stato raggiunto al passo n-esimo e successivamente applicare la condizione di Markov. Poiché $j \neq 1$ abbiamo

$$\begin{aligned}
\rho_j &= \sum_{n=1}^{+\infty} \Pr_1(X_n = j,\ \tau_1 > n) \ = \ \sum_{n=1}^{+\infty} \Pr_1(X_n = j,\ \tau_1 > n-1) \\
&= \sum_{n=1}^{+\infty} \sum_{i=1}^{k} \Pr_1(X_n = j,\ X_{n-1} = i,\ \tau_1 > n-1) \\
&= \sum_{n=1}^{+\infty} \sum_{i=1}^{k} \Pr_1(X_n = j \mid X_{n-1} = i,\ \tau_1 > n-1)\ \Pr_1(X_{n-1} = i,\ \tau_1 > n-1)
\end{aligned}$$

Considera il termine generale dell'ultima sommatoria: se $i \neq 1$ e $n > 1$ (oppure $i = 1 = n$) il primo fattore coincide con $\Pr_1(X_n = j \mid X_{n-1} = i) = p(i,j)$; se invece $i = 1$ e $n > 1$ oppure $i \neq 1$ e $n = 1$ allora il secondo fattore è nullo, cioé $\Pr_1(X_{n-1} = i,\ \tau_1 > n-1) = 0$. Di conseguenza possiamo scrivere

$$\begin{aligned}
\rho_j &= \sum_{n=1}^{+\infty} \sum_{i=1}^{k} p(i,j)\ \Pr_1(X_{n-1} = i,\ \tau_1 > n-1) \\
&= \sum_{i=1}^{k} p(i,j)\ \sum_{n=1}^{+\infty} \Pr_1(X_{n-1} = i,\ \tau_1 > n-1) \\
&= \sum_{i=1}^{k} p(i,j)\ \rho_i \ = \ (\rho'\, P)_j
\end{aligned}$$

Consideriamo ora ρ_1 che per l'osservazione iniziale coincide con 1. Poiché lo stato 1 è ricorrente abbiamo

$$\begin{aligned}
\rho_1 &= \sum_{n=1}^{+\infty} \Pr_1(\tau_1 = n) \\
&= \sum_{n=1}^{+\infty} \sum_{i=1}^{k} \Pr_1(X_{n-1} = i,\ \tau_1 = n) \qquad (\text{poiché } \{\tau_1 = n\} = \{X_n = 1,\ \tau_1 > n-1\}) \\
&= \sum_{n=1}^{+\infty} \sum_{i=1}^{k} \Pr_1(X_n = 1 \mid X_{n-1} = i,\ \tau_1 > n-1)\ \Pr_1(X_{n-1} = i,\ \tau_1 > n-1) \\
&= \sum_{n=1}^{+\infty} \sum_{i=1}^{k} p(i,1)\ \Pr_1(X_{n-1} = i,\ \tau_1 > n-1) \\
&= \sum_{i=1}^{k} p(i,1) \sum_{n=1}^{+\infty} \Pr_1(X_{n-1} = i,\ \tau_1 > n-1) \ = \ \sum_{i=1}^{k} p(i,1)\ \rho_i \ = \ (\rho'\, P)_1
\end{aligned}$$

Abbiamo quindi provato che $\rho' = (\rho'\, P)$ e quindi $\pi' = (\pi'\, P)$. □

Corollario 3.3 *Nell'ipotesi della proposizione precedente, se P è una matrice irriducibile allora π è strettamente positivo.*

Dimostrazione. Per la irriducibilità di P, nel grafo associato alla catena di Markov, per ogni nodo $i \neq 1$ esiste un cammino semplice da 1 a i e quindi per qualche $n > 0$, $\Pr_1(X_n = i,\ \tau_1 > n) > 0$. Questo prova che ogni ρ_i è positivo e di conseguenza risulta $\pi > 0$. □

Osservazione 5 *Se i è uno stato ricorrente diverso da 1, il vettore π può essere costruito a partire da i invece che da 1. La proposizione 3.2 garantisce che anche in questo caso π è una distribuzioni stazionaria per la stessa catena. In particolare la sua componente i-esima risulta $\pi = (E_i(\tau_i))^{-1}$. Nelle prossime sezioni dimostreremo che se i due stati appartengono alla stessa classe le due distribuzioni stazionarie coincidono.*

3.2 Catene di Markov primitive

In questa sezione mostriamo che le catene di Markov con matrice di transizione primitiva sono ergodiche e che la distribuzione limite coincide con l'unica distribuzione stazionaria. Ricordiamo innanzitutto la definizione della norma ℓ_1 su $\mathbb{R}^k$ e della relativa distanza: dato un intero $k > 0$, per ogni $v, w \in \mathbb{R}^k$ poniamo

$$\begin{aligned} \| v \|_1 &= \sum_{i=1}^{k} |v_i| \\ \mathrm{dist}(v,w) &= \| v - w \|_1 = \sum_{i=1}^{k} |v_i - w_i| \end{aligned}$$

Nota che $\mathrm{dist}(v,w) \leq 2$, per ogni coppia di vettori stocastici $v, w \in \mathbb{R}^k$.

Vogliamo ora dimostrare che per ogni matrice stocastica primitiva P di dimensione k e ogni coppia di vettori stocastici $\mu, \nu \in \mathbb{R}^k$, vale

$$\lim_{n \to +\infty} \| \mu' P^n - \nu' P^n \|_1 = 0$$

A tale scopo, consideriamo due catene di Markov $\{X_n\}$ e $\{Y_n\}$ con matrice di transizione P e distribuzioni iniziali μ e ν, rispettivamente. Supponiamo che $\{X_n\}$ e $\{Y_n\}$ siano definite sullo stesso spazio di probabilità e siano indipendenti, nel senso che ogni X_i è indipendente da ogni Y_j. Si può immaginare una simile coppia facendo evolvere le due catene in parallelo su due traiettorie distinte (quindi lo spazio di probabilità comune è quello delle coppie ordinate di traiettorie, che possiamo supporre di sviluppare concorrentemente). Definiamo allora la variabile aleatoria T che rappresenta l'eventuale istante di incontro delle due catene:

$$T = \min\{n \in \mathbb{N} \mid X_n = Y_n\}$$

Anche in questo caso, T assume valori in $\mathbb{N} \cup \{+\infty\}$, dove poniamo $T = +\infty$ se non esiste alcun $n \in \mathbb{N}$ tale che $X_n = Y_n$. Mostriamo innanzitutto che $\Pr(T = +\infty) = 0$.

Lemma 3.4 *Per qualche ε tale che $0 < \varepsilon < 1$, abbiamo*

$$Pr(T > n) = O(\varepsilon^n)$$

Dimostrazione. Poiché P è primitiva, esiste un minimo $m \in \mathbb{N}$ tale che $P^m > 0$. Possiamo allora definire

$$\alpha = \min\{p^{(m)}(i,j) \mid i,j = 1,2,\ldots,k\}$$

Essendo P^m una matrice stocastica, abbiamo $0 < \alpha \leq 1/k$. Inoltre, possiamo fornire un limite inferiore alla probabilità che le due catene si incontrino entro il passo m-esimo:

$$\begin{aligned}
\Pr(T \leq m) &\geq \Pr(X_m = Y_m) \\
&\geq \Pr(X_m = 1, Y_m = 1) \\
&= \Pr(X_m = 1) \cdot \Pr(Y_m = 1) \\
&= (\mu' P^m)_1 \cdot (\nu' P^m)_1 \;\geq\; \left(\alpha \sum_{i=1}^{k} \mu_i\right) \cdot \left(\alpha \sum_{i=1}^{k} \nu_i\right) \;=\; \alpha^2
\end{aligned}$$

Di conseguenza

$$\Pr(T > m) \leq (1 - \alpha^2)$$

In modo analogo si prova che $\Pr(T > 2m \mid T > m) \leq (1 - \alpha^2)$. Infatti, per ogni evento A occorso entro i primi m passi, abbiamo

$$\begin{aligned}
\Pr(X_{2m} = Y_{2m} \mid A) &= \sum_{i,j=1}^{k} \Pr(X_{2m} = Y_{2m} \mid X_m = i,\, Y_m = j,\, A) \Pr(X_m = i,\, Y_m = j \mid A) \\
&= \sum_{i,j=1}^{k} \Pr(X_{2m} = Y_{2m} \mid X_m = i,\, Y_m = j) \Pr(X_m = i,\, Y_m = j \mid A) \\
&\geq \sum_{i,j=1}^{k} \Pr(X_{2m} = Y_{2m} = 1 \mid X_m = i,\, Y_m = j) \Pr(X_m = i,\, Y_m = j \mid A) \\
&= \sum_{i,j=1}^{k} p^{(m)}(i,1) p^{(m)}(j,1) \Pr(X_m = i,\, Y_m = j \mid A) \\
&\geq \alpha^2 \sum_{i,j=1}^{k} \Pr(X_m = i,\, Y_m = j \mid A) \;=\; \alpha^2
\end{aligned}$$

Di conseguenza

$$\begin{aligned}
\Pr(T > 2m) &= \Pr(T > 2m \mid T > m) \Pr(T > m) \\
&\leq \Pr(T > 2m \mid T > m)\, (1 - \alpha^2) \\
&\leq \Pr(X_{2m} \neq Y_{2m} \mid T > m)\, (1 - \alpha^2) \\
&= [1 - \Pr(X_{2m} = Y_{2m} \mid T > m)]\, (1 - \alpha^2) \;\leq\; (1 - \alpha^2)^2
\end{aligned}$$

Iterando il ragionamento, per ogni intero $\ell > 1$ otteniamo

$$\begin{aligned}
\Pr(T > \ell m) &= \Pr(T > \ell m \mid T > (\ell - 1)m) \Pr(T > (\ell - 1)m) \\
&\leq [1 - \Pr(X_{\ell m} = Y_{\ell m} \mid T > (\ell - 1)m)]\, (1 - \alpha^2)^{(\ell - 1)} \;\leq\; (1 - \alpha^2)^{\ell}
\end{aligned}$$

Per ogni $n \in \mathbb{N}$ abbastanza grande possiamo porre $n = sm + r$, con $r \in \{0, 1, \ldots, m-1\}$ e quindi per l'equazione precedente otteniamo

$$\begin{aligned}
\Pr(T > n) &\leq \Pr(T > sm) \;\leq\; (1 - \alpha^2)^s \\
&= (1 - \alpha^2)^{\frac{n-r}{m}} \;=\; O((1 - \alpha^2)^{1/m})^n
\end{aligned}$$

che prova l'enunciato poiché $0 < (1 - \alpha^2)^{1/m} < 1$. □

Corollario 3.5
$$Pr(T=+\infty) \;=\; \lim_{n\to+\infty} Pr(T>n) \;=\; 0$$

Dimostrazione. La prima uguaglianza è ovvia perchè $\{T=+\infty\}=\bigcap_{n\geq 1}\{T>n\}$ e $\{T>n\}\supseteq\{T>n-1\}$. La seconda segue invece dal lemma 3.4. □

Abbiamo dunque provato che con probabilità 1 le due catene si incontrano. Usiamo ora questo fatto per dimostrare che le funzioni probabilità di X_n e Y_n si avvicinano, nel senso che tendono a definire la stessa partizione dell'intervallo $[0,1]$.

Proposizione 3.6 *Per ogni matrice stocastica primitiva P di dimensione k e ogni coppia di vettori stocastici $\mu,\nu\in\mathbb{R}^k$, vale la relazione*
$$\lim_{n\to+\infty} \| \mu' P^n - \nu' P^n \|_1 \;=\; 0$$

Dimostrazione. Proseguendo il ragionamento avviato nel lemma 3.4, consideriamo una nuova catena di Markov $\{Z_n\}$ tale che: $Z_0=X_0$, Z_n coincida con X_n per tutti gli $n\leq T$, Z_n coincida con Y_n per ogni $n>T$. In altre parole, Z_n è definita sulla traiettoria di X_n fino al primo istante i tale che $X_i=Y_i$; non appena si verifica questo evento, la variabile Z_n passa sulla traiettoria di Y_n. Nota che la distribuzione di Z_n coincide con quella di X_n, ovvero
$$\Pr(Z_n=i) \;=\; (\mu' P^n)_i \qquad\qquad (\forall\, i=1,2,\ldots,k)$$
Allora, per ogni $i=1,2,\ldots,k$ possiamo scrivere
$$\begin{aligned}(\mu' P^n)_i-(\nu' P^n)_i \;&=\; \Pr(Z_n=i)-\Pr(Y_n=i)\\ &\leq\; \Pr(Z_n=i)-\Pr(Z_n=i,\,Y_n=i)\\ &\leq\; \Pr(Z_n=i,\,Y_n\neq i) \;\leq\; \Pr(Z_n\neq Y_n) \;=\; \Pr(T>n)\end{aligned}$$
In maniera analoga, si prova che
$$(\nu' P^n)_i-(\mu' P^n)_i \;\leq\; \Pr(T>n)$$
e quindi per il lemma 3.4 otteniamo
$$|(\mu' P^n)_i-(\nu' P^n)_i| \;\leq\; 2\;\Pr(T>n) \;=\; \mathrm{O}(\varepsilon^n) \qquad (0<\varepsilon<1)$$
Sommando ora su tutti gli stati $i=1,2,\ldots,k$ si ricava
$$\| \mu' P^n - \nu' P^n \|_1 \;=\; \mathrm{O}(\varepsilon^n) \qquad (0<\varepsilon<1)$$
□

Teorema 3.7 *Sia $\{X_n\}$ una catena di Markov sull'insieme di stati $\{1,2,\ldots,k\}$, con matrice di transizione primitiva P e distribuzione iniziale μ. Allora, tale catena ammette un'unica distribuzione stazionaria $\pi^*=(\pi_1^*,\pi_2^*,\ldots,\pi_k^*)$. Inoltre, per ogni $i=1,2,\ldots,k$, valgono le relazioni*
$$\pi_i^* \;=\; \frac{1}{E_i(\tau_i)} \;=\; \frac{1}{\sum_{n\geq 1} n f^{(n)}(i,i)}$$
e
$$\lim_{n\to+\infty} Pr_\mu(X_n=i) \;=\; \pi_i^*$$

Dimostrazione. Poiché P è irriducibile, il vettore π definito in (3.1) è una distribuzione stazionaria. Ne segue che $\pi' P^n = \pi'$ per ogni $n \geq 1$ e quindi, per la proposizione 3.6

$$\lim_{n\to+\infty} \| \mu' P^n - \pi' \|_1 = 0 \tag{3.2}$$

Questo prova anche l'unicità della distribuzione stazionaria per la catena $\{X_n\}$: infatti, ogni altra distribuzione stazionaria η deve soddisfare $\eta' P^n = \eta'$ e $\| \eta' P^n - \pi' \|_1 \to 0$; quindi $\| \eta' - \pi' \|_1 \to 0$ e di conseguenza $\eta = \pi$. In particolare, se costruiamo una distribuzione stazionaria ν come nella sezione precedente partendo da uno stato qualsiasi i (invece che dallo stato 1), otteniamo $\nu_i = 1/E_i(\tau_i)$. Per l'unicità abbiamo $\nu = \pi$ e questo prova $\pi = \pi^*$. Quindi π^* coincide con l'unica distribuzione stazionaria. La seconda equazione dell'enunciato è allora conseguenza della relazione (3.2). □

3.3 Matrici primitive e contrazioni

In questa sezione presentiamo un diverso approccio allo studio delle proprietà di convergenza delle catene di Markov primitive. Più precisamente mostriamo come si possa provare l'ergodicità della catena usando le proprietà degli spazi metrici. A tale scopo iniziamo ricordando il noto teorema del punto fisso di Banach.

Teorema 3.8 *Sia M uno spazio metrico completo con distanza $d : M \times M \to \mathbb{R}_+$ e sia la funzione $f : M \to M$ una contrazione su M, ovvero per qualche β reale, $0 < \beta < 1$,*

$$d(f(x), f(y)) \leq \beta d(x,y)$$

per ogni $x, y \in M$. Allora esiste un unico elemento $z \in M$ tale che

$$f(z) = z$$

e inoltre, per ogni $x \in M$

$$\lim_{n\to+\infty} f^{(n)}(x) = z$$

dove $f^{(n)}$ rappresenta la composizione di f con sè stessa n volte.

Cerchiamo di applicare il teorema precedente al nostro caso. A questo scopo consideriamo, per ogni intero $k > 1$, l'insieme M_k dei vettori stocastici di dimensione k:

$$M_k = \{v \in \mathbb{R}_+^k \mid \sum_{i=1}^k v_i = 1\}$$

Tale insieme forma uno spazio metrico chiuso e completo rispetto alla distanza ℓ_1, data da

$$\| v - w \|_1 = \sum_{i=1}^k |v_i - w_i| \qquad (\forall\, v, w \in M_k)$$

Osserviamo inoltre che per ogni $x \in \mathbb{R}^k$ tale che $\sum_{i=1}^k x_i = 0$ si ha

$$\sum_{i\,:\,x_i \geq 0} x_i = \sum_{j\,:\,x_j < 0} |x_j| = \frac{\| x \|_1}{2} \tag{3.3}$$

Lemma 3.9 *Sia A una matrice stocastica positiva di dimensione k (quindi $A_{ij} > 0$ per ogni i, j), e sia $\alpha = \min\{A_{ij} \mid i, j = 1, 2, \ldots, k\}$. Allora, per ogni $x \in \mathbb{R}^k$ tale che $\sum_{i=1}^k x_i = 0$, abbiamo*

$$\| x'A \|_1 \;\leq\; (1 - \alpha k) \| x \|_1$$

Nota che $0 \leq (1 - \alpha k) < 1$.

Dimostrazione. Proviamo innazitutto che per ogni $j = 1, 2, \ldots, k$, vale la disuguaglianza

$$\left| \sum_{i=1}^k x_i A_{ij} \right| \;\leq\; \sum_{i=1}^k |x_i| A_{ij} - 2\alpha \sum_{i\,:\,x_i < 0} |x_i|$$

Infatti, se $\sum_{i=1}^k x_i A_{ij} \geq 0$ allora possiamo scrivere

$$\begin{aligned}
\left| \sum_{i=1}^k x_i A_{ij} \right| &= \sum_{i\,:\,x_i \geq 0} x_i A_{ij} - \sum_{i\,:\,x_i < 0} |x_i| A_{ij} \\
&= \sum_{i=1}^k |x_i| A_{ij} - 2 \sum_{i\,:\,x_i < 0} |x_i| A_{ij} \;\leq\; \sum_{i=1}^k |x_i| A_{ij} - 2\alpha \sum_{i\,:\,x_i < 0} |x_i|
\end{aligned}$$

Se invece $\sum_{i=1}^k x_i A_{ij} < 0$, con un ragionamento analogo otteniamo

$$\left| \sum_{i=1}^k x_i A_{ij} \right| = \sum_{i=1}^k |x_i| A_{ij} - 2 \sum_{i\,:\,x_i \geq 0} x_i A_{ij} \;\leq\; \sum_{i=1}^k |x_i| A_{ij} - 2\alpha \sum_{i\,:\,x_i < 0} |x_i|$$

dove la disuguaglianza segue dalla relazione (3.3).

L'enunciato è quindi provato dalle seguenti disuguaglianze:

$$\begin{aligned}
\| x'A \|_1 &= \sum_{j=1}^k \left| \sum_{i=1}^k x_i A_{ij} \right| \;\leq\; \sum_{j=1}^k \left(\sum_{i=1}^k |x_i| A_{ij} - 2\alpha \sum_{i\,:\,x_i < 0} |x_i| \right) = \\
&= \sum_{i=1}^k |x_i| - 2\alpha k \sum_{i\,:\,x_i < 0} |x_i| = (1 - \alpha k) \sum_{i=1}^k |x_i| = (1 - \alpha k) \| x \|_1
\end{aligned}$$

□

Di conseguenza, se A è definita come nel lemma precedente, la funzione $x \longrightarrow x'A$ ($x \in M_k$) definisce una contrazione su M_k ed ammette quindi un unico punto fisso. Siamo così in grado di provare nuovamente il teorema 3.7.

Teorema 3.10 *Sia $\{X_n\}$ una catena di Markov sull'insieme di stati $\{1, 2, \ldots, k\}$, con matrice di transizione primitiva P e distribuzione iniziale μ. Allora, tale catena ammette un'unica distribuzione stazionaria $\pi^* = (\pi_1^*, \pi_2^*, \ldots, \pi_k^*)$. Inoltre, per ogni $i = 1, 2, \ldots, k$, abbiamo*

$$\pi_i^* = \frac{1}{E_i(\tau_i)} = \frac{1}{\sum_{n \geq 1} n f^{(n)}(i,i)}$$

e

$$\lim_{n \to +\infty} Pr_\mu(X_n = i) = \pi_i^*$$

Dimostrazione. Poiché P è primitiva esiste un intero $m > 0$ tale che $P^m > 0$. Per il lemma 3.9 la funzione $f : M_k \to M_k$ tale che

$$f(v) \;=\; v'P^m \qquad\qquad (\forall\, v \in M_k)$$

è una contrazione su M_k. Quindi, per il teorema di Banach esiste un unico $z \in M_k$ tale che $z' = z'P^m$ e inoltre

$$\lim_{u\to+\infty} \mu'(P^m)^u \;=\; z' \qquad\qquad (\forall\, \mu \in M_k)$$

Considera ora il vettore π definito nella (3.1). Poiché anche π verifica $\pi' = \pi'P^m$, otteniamo $\pi = z$ e quindi

$$\lim_{u\to+\infty} \mu'(P^m)^u \;=\; \pi' \qquad\qquad (\forall\, \mu \in M_k) \tag{3.4}$$

La stessa cosa vale per ogni distribuzione stazionaria costruita come nella sezione 3.1 a partire da un generico stato i invece che da 1. Tali distribuzioni sono tutte identiche e quindi $\pi_i = 1/E_i(\tau_i)$ $(= \pi_i^*)$ per ogni $i = 1, 2, \ldots, k$.

Infine, per un generico $n \in \mathbb{N}$ abbastanza grande, siano $r, q \in \mathbb{N}$ tali che $n = mq + r$ e $0 \le r < m$. Al crescere di n, r assume un numero finito di valori, Quindi dalla relazione (3.4) si deduce che anche $\mu'P^n = \mu'P^r(P^m)^q$ converge a π^* per $n \to +\infty$ e questo prova l'enunciato. □

3.4 Catene di Markov irriducibili periodiche

Nelle sezioni precedenti abbiamo visto che le catene di Markov primitive sono ergodiche. Invece, per le catene di Markov irriducibili di periodo maggiore di 1 questa proprietà non è più vera. In questo caso infatti la sequenza $\{p^{(n)}(i,j)\}_n$ in generale non ammette limite.

Esempio 3.1 Considera una catena formata da due soli stati $\{1,2\}$ con matrice di transizione

$$P \;=\; \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Chiaramente P è irriducibile di periodo 2 e si verifica subito che $p^{(n)}(1,2) = 1$ se n è dispari, mentre $p^{(n)}(1,2) = 0$ per n pari. Quindi, al crescere di n, $\{p^{(n)}(1,2)\}_n$ non ammette limite e lo stesso vale per ogni altra coppia di stati.

Si può però dimostrare che anche nel caso irriducibile periodico la distribuzione stazionaria è unica e coincide con il vettore π^* definito nel teorema 3.7. Valgono inoltre proprietà di limite per sottosequenze di $\{p^{(n)}(i,j)\}$ e si può provare una forma di convergenza più debole, relativa cioé alla sequenza delle somme $\{\frac{1}{n}\sum_{k=0}^{n} p^{(k)}(i,j)\}_n$ (intuitivamente, una sorta di media dei valori $p^{(n)}(i,j)$, al variare di n).

Prima di illustrare questi risultati ricordiamo alcune proprietà delle matrici irriducibili presentate nel capitolo 1 adattandole alla nostra situazione.

Data una matrice stocastica irriducibile $P = [p(u,v)]$ di periodo $d > 1$ e dimensione k, esistono d famiglie di indici $C_0, C_1, \ldots, C_{d-1}$ $(C_d = C_0)$ che formano una partizione dell'insieme $\{1, 2, \ldots, k\}$ e che rappresentano le "classi di periodicità" di P. Esse verificano in particolare le seguenti proprietà.

1) Se $u \in C_i$ e $p(u,v) > 0$ allora $v \in C_{i+1}$.

2) Due stati u, v appartengono allo stesso insieme C_i se e solo se $p^{(nd)}(u,v) > 0$ per qualche $n \in \mathbb{N}$.

3) Per ogni $u \in C_i$ e ogni $v \in C_j$, posto $r \in \{0,1,\ldots,d-1\}$ tale che $r \equiv i - j \pmod d$, allora

$$p^{(n)}(u,v) > 0 \ \text{ implica } \ n \equiv r \pmod d$$

4) La restrizione di P^d agli indici in C_i forma una matrice stocastica primitiva. Quindi, permutando l'ordine di righe e colonne, P^d può essere trasformata in una forma diagonale a blocchi

$$(P^d)^* \ = \ \begin{bmatrix} U_0 & 0 & \cdots & 0 \\ 0 & U_1 & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & U_{d-1} \end{bmatrix}$$

nella quale le matrici U_i sono disposte sulla diagonale principale, sono quadrate e coincidono con la restrizione di P^d agli indici in C_i (risultano quindi primitive); viceversa, ogni altro elemento della matrice è nullo.

5) Analoghe proprietà di periodicità valgono per i coefficienti $f^{(n)}(u,v)$ introdotti nella sezione 2.4. Poiché $f^{(n)}(u,v) > 0$ implica $p^{(n)}(u,v) > 0$ abbiamo

$$f^{(n)}(u,u) = 0 \qquad \text{per ogni } \ n \not\equiv 0 \pmod d, \ \text{ e ogni } \ u. \tag{3.5}$$

Per lo stesso motivo, se $u \in C_i$, $v \in C_j$ e $r \in \{0,1,\ldots,d-1\}$ tale che $r \equiv i - j \pmod d$, allora

$$f^{(n)}(u,v) = 0 \qquad \text{per ogni } \ n \not\equiv r \pmod d \tag{3.6}$$

e quindi, essendo P irriducibile,

$$1 = f(u,v) = \sum_{n\geq 1} f^{(n)}(u,v) = \sum_{n\geq 1} f^{(nd+r)}(u,v) \tag{3.7}$$

Ricordiamo infine che, essendo ogni stato $u \in \{1,2,\ldots,k\}$ ricorrente, il valor medio di τ_u è sempre ben definito:

$$E_u(\tau_u) \ = \ \sum_{n\geq 1} n f^{(n)}(u,u) \ < +\infty$$

Per una tale matrice P possiamo allora definire il vettore $\pi^* = (\pi_1^*, \ldots, \pi_k^*)$ tale che

$$\pi_u^* \ = \ \frac{1}{E_u(\tau_u)} \ = \ \frac{1}{\sum_{n\geq 1} n f^{(n)}(u,u)} \qquad (\forall\, u \in \{1,2,\ldots,k\})$$

Proposizione 3.11 *Se P è una matrice stocastica irriducibile di dimensione k e periodo d allora*

$$\lim_{n\to+\infty} p^{(nd)}(u,u) \ = \ d\pi_u^* \qquad (\forall\, u \in \{1,2,\ldots,k\})$$

Nota che per la proprietà 3) data sopra, $p^{(n)}(u,u) = 0$ per ogni $n \not\equiv 0 \pmod d$. Quindi, se $d > 1$ la proposizione implica che $p^{(n)}(u,u)$ non ammette limite al crescere di n. Ne segue che una catena di Markov irriducibile periodica non può essere ergodica.

Dimostrazione. Sia u un indice in $\{1,2,\ldots,k\}$. Innanzitutto osserviamo che per l'equazione (3.5) il coefficiente π_u^* verifica $\pi_u^* = 1/\sum_{n\geq 1} nd\, f^{(nd)}(u,u)$.

Denotata con C la classe di periodicità di u, per ogni $n \in \mathbb{N}$ e ogni $x,y \in C$, definiamo

$$h^{(n)}(x,y) = p^{(nd)}(x,y)\ , \qquad g^{(n)}(x,y) = f^{(nd)}(x,y)$$

Chiaramente la matrice $H = [h^{(1)}(x,y)]_{x,y\in C}$ è la restrizione di P^d alle componenti in C e quindi è stocastica primitiva. Analogamente, i valori $g^{(n)}(x,y)$ rappresentano le probabilità di prima entrata da x a y in n passi in una catena di Markov con spazio degli stati C e matrice di transizione H. Essendo H primitiva, per il teorema 3.7, abbiamo

$$\lim_{n\to+\infty} h^{(n)}(u,u) \;=\; \frac{1}{\sum_{n\geq 1} n\, g^{(n)}(u,u)}$$

ovvero

$$\lim_{n\to+\infty} p^{(nd)}(u,u) \;=\; \frac{1}{\sum_{n\geq 1} n\, f^{(nd)}(u,u)} \;=\; d\pi_u^*$$

□

Proposizione 3.12 *Sia P una matrice stocastica irriducibile di periodo $d > 1$ e siano C_i e C_j due classi di periodicità. Allora, per ogni $u \in C_i$ e ogni $v \in C_j$ abbiamo*

$$\lim_{n\to+\infty} p^{(nd+r)}(u,v) \;=\; d\,\pi_v^*$$

dove $r \equiv i-j \pmod d$, $r \in \{0,1,\ldots,d-1\}$.

Omettiamo la prova di questa proposizione. Segnaliamo solo che la dimostrazione si ottiene essenzialmente in due passi: prima si considera la relazione tra i coefficienti $p^{(n)}(i,j)$ e gli $f^{(n)}(i,j)$ data nella proposizione 2.2, che nel nostro caso diventa

$$p^{(nd+r)}(u,v) = \sum_{j=0}^{nd+r} f^{(j)}(u,v)p^{(nd+r-j)}(v,v) = \sum_{j=0}^{n} f^{(jd+r)}(u,v)p^{(n-j)d}(v,v)\ ;$$

quindi si applica la proposizione 3.11 al termine destro della sommatoria e si ricorda che

$$\lim_{n\to+\infty} f^{(n)}(u,v) = 0$$

Osserviamo infine che nelle stesse ipotesi precedenti $p^{(n)}(u,v) = 0$ per ogni $n \not\equiv r \pmod d$. Si può così dedurre il seguente risultato.

Corollario 3.13 *Se P è una matrice stocastica irriducibile allora, per ogni coppia di indici i,j abbiamo*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{\ell=0}^{n} p^{(\ell)}(i,j) \;=\; \pi_j^*$$

Per quanto riguarda le distribuzioni stazionarie delle catene irriducibili, sappiamo già dalla sezione 3.1 che ogni matrice stocastica e irriducibile P ammette una distribuzione stazionaria π definita in (3.1). Il prossimo enunciato afferma che π coincide con π^* e che questa è l'unica distribuzione stazionaria di P.

Teorema 3.14 *Ogni catena di Markov irriducibile ammette un'unica distribuzione stazionaria che coincide con il vettore* π^*.

Dimostrazione. Sia $P = [p(u,v)]$ la matrice di transizione della catena e sia μ una sua distribuzione stazionaria. Poiché P è irriducibile, per l'osservazione precedente sappiamo che ne esiste almeno una. Denotiamo con d il periodo di P e con $C_0, C_1, \cdots, C_{d-1}$ le classi di periodicità corrispondenti $(C_d = C_0)$. Allora

$$\begin{aligned}\sum_{v\in C_{i+1}} \mu_v &= \sum_{v\in C_{i+1}} \sum_{u\in C_i} \mu_u p(u,v) \\ &= \sum_{u\in C_i} \mu_u \sum_{v\in C_{i+1}} p(u,v) = \sum_{u\in C_i} \mu_u\end{aligned}$$

Questo prova che $\sum_{u\in C_i} \mu_u = 1/d$ per ogni $i = 0, 1 \ldots, d-1$.

Inoltre, essendo μ stazionario, $\mu' P^{nd} = \mu'$ per ogni $n \in \mathbb{N}$ e per ogni stato v abbiamo

$$(\mu' P^{nd})_v = \mu_v$$

Applicando la proposizione 3.12 al termine sinistro, posto $v \in C_i$, abbiamo

$$(\mu' P^{nd})_v = \sum_{u\in C_i} \mu_u p^{(nd)}(u,v) \longrightarrow \sum_{u\in C_i} \mu_u \cdot d\, \pi^*_v = \pi^*_v$$

e per l'equazione precedente deve essere $\mu_v = \pi^*_v$. □

3.5 Catene di Markov riducibili

È ora naturale chiedersi quale sia il comportamento asintotico di $\{p^{(n)}(i,j)\}_n$ per le catene di Markov riducibili, nelle quali cioé il grafo di connessione contiene più componenti fortemente connesse. In questo caso non esiste una risposta univoca e il comportamento dipende dalla relazione di connessione tra le varie componenti e dalla periodicità dei loro elementi. Si può però provare che se vi è una sola classe essenziale e quest'ultima è aperiodica allora la catena è ergodica. Viceversa, se vi sono più classi essenziali, oppure se l'unica classe essenziale è periodica, allora la catena non può essere ergodica.

Per illustrare questi risultati, consideriamo come sempre una catena di Markov $\{X_n\}$ su un insieme finito di stati S, con matrice di transizione $P = [p(i,j)]_{i,j\in S}$.

Osserviamo innanzitutto che per il corollario 2.10 se $j \in S$ è uno stato transiente allora per ogni $i \in S$

$$\lim_{n\to+\infty} p^{(n)}(i,j) = 0$$

Rimane quindi da indagare il comportamento di $\{p^{(n)}(i,j)\}_n$ nel caso di j ricorrente.

Proposizione 3.15 *Se la catena di Markov* $\{X_n\}$ *possiede una sola classe essenziale* C *e quest'ultima è aperiodica allora, per ogni* $j \in C$ *e ogni* $i \in S$,

$$\lim_{n\to+\infty} p^{(n)}(i,j) = \frac{1}{\sum_{n\geq 1} n f^{(n)}(j,j)}$$

Per l'osservazione precedente la catena è quindi ergodica.

Dimostrazione. Proviamo innanzitutto che $f(i,j) = 1$. Se $i \in C$ la proprietà segue dalla proposizione 2.7. Supponiamo che i appartenga all'insieme T degli stati transienti e denotiamo con σ_T il tempo di permanenza della catena in T, cioé $\sigma_T = \max\{n \in \mathbb{N} \mid X_n \in T\}$. Per la proposizione 2.11, sappiamo che $\Pr_i(\sigma_T = +\infty) = 0$ e quindi, essendo $C = S \backslash T$, abbiamo

$$1 \;=\; \sum_{\ell=0}^{+\infty} \Pr_i(\sigma_T = \ell) \;=\; \sum_{\ell=0}^{+\infty} \sum_{u \in C} \Pr_i(\sigma_T = \ell, X_{\ell+1} = u) \tag{3.8}$$

Per valutare $f(i,j)$ possiamo allora condizionare sull'evento $\{\sigma_T = \ell, X_{\ell+1} = u\}$:

$$\begin{aligned}
f(i,j) &= \Pr_i(\exists\, n \;:\; \tau_j = n) \;=\; \sum_{\ell=0}^{+\infty} \sum_{u \in C} \Pr_i(\exists n \;:\; \tau_j = n,\; \sigma_T = \ell, X_{\ell+1} = u) \;= \\
&= \sum_{\ell=0}^{+\infty} \sum_{u \in C} \Pr_i(\exists n > \ell \;:\; \tau_j = n,\; \sigma_T = \ell, X_{\ell+1} = u) \;= \\
&= \sum_{\ell=0}^{+\infty} \sum_{u \in C} \Pr_i(\exists n > \ell \;:\; \tau_j = n \mid\; \sigma_T = \ell, X_{\ell+1} = u)\, \Pr_i(\sigma_T = \ell, X_{\ell+1} = u) \;= \\
&= \sum_{\ell=0}^{+\infty} \Pr_i(\sigma_T = \ell, X_{\ell+1} = j) + \sum_{u \in C, u \neq j} \Pr_u(\exists n > 0 \;:\; \tau_j = n)\, \Pr_i(\sigma_T = \ell, X_{\ell+1} = u) \;= \\
&= \sum_{\ell=0}^{+\infty} \sum_{u \in C} f(u,j)\, \Pr_i(\sigma_T = \ell, X_{\ell+1} = u)
\end{aligned}$$

Per la proposizione 2.7 $f(u,j) = 1$ per ogni $u \in C$ e quindi dalla (3.8) ricaviamo

$$f(i,j) = \sum_{\ell=0}^{+\infty} \sum_{u \in C} \Pr_i(\sigma_T = \ell, X_{\ell+1} = u) = 1$$

Poiché C è una classe essenziale e aperiodica, la restrizione di P agli indici in C forma una matrice primitiva. Possiamo quindi applicare il teorema 3.7, ottenendo

$$\lim_{n \to +\infty} p^{(n)}(j,j) \;=\; \pi_j^* \;=\; \frac{1}{\sum_{n \geq 1} n f^{(n)}(j,j)}$$

Questo significa che, per ogni $\varepsilon > 0$, esiste $m_\varepsilon \in \mathbb{N}$ tale che per ogni intero $n > m_\varepsilon$, $|p^{(n)}(j,j) - \pi_j^*| \leq \varepsilon$. Di conseguenza, per n abbastanza grande, la nota equazione (2.1) che lega i coefficienti $p^{(n)}(j,j)$ e $f^{(n)}(j,j)$, può essere riscritta nella forma seguente:

$$\begin{aligned}
p^{(n)}(i,j) &= \sum_{k=1}^{n} f^{(k)}(i,j) p^{(n-k)}(j,j) \\
&= \sum_{k=1}^{n-m_\varepsilon-1} f^{(k)}(i,j) p^{(n-k)}(j,j) \;+\; \sum_{k=n-m_\varepsilon}^{n} f^{(k)}(i,j) p^{(n-k)}(j,j) \\
&= \sum_{k=m_\varepsilon+1}^{n-1} f^{(n-k)}(i,j) p^{(k)}(j,j) \;+\; \sum_{k=0}^{m_\varepsilon} f^{(n-k)}(i,j) p^{(k)}(j,j)
\end{aligned} \tag{3.9}$$

Poiché i termini della seconda sommatoria sono positivi o nulli, per la scelta di m_ε abbiamo

$$p^{(n)}(i,j) \;\geq\; (\pi_j^* - \varepsilon) \sum_{k=m_\varepsilon+1}^{n-1} f^{(n-k)}(i,j)$$

Per ogni $\varepsilon > 0$, la sommatoria nel termine destro tende a $f(i,j) = 1$ per $n \to +\infty$ e quindi

$$\liminf_{n\to+\infty} p^{(n)}(i,j) \ \geq \ (\pi_j^* - \varepsilon) \tag{3.10}$$

Un limite analogo vale per l'estremo superiore e questo conclude la prova; infatti, dall'equazione (3.9), sempre per la scelta di m_ε otteniamo la disuguaglianza

$$p^{(n)}(i,j) \ \leq \ (\pi_j^* + \varepsilon) + \sum_{k=0}^{m_\varepsilon} f^{(n-k)}(i,j)$$

e poiché $f^{(n)}(i,j) \to 0$ per $n \to +\infty$, si ricava

$$\limsup_{n\to+\infty} p^{(n)}(i,j) \ \leq \ (\pi_j^* + \varepsilon) \tag{3.11}$$

Per l'arbitrarietà di ε, le relazioni (3.10) e (3.11) implicano

$$\lim_{n\to+\infty} p^{(n)}(i,j) \ = \ \pi_j^* \tag{3.12}$$

□

Osserviamo che una catena di Markov finita con due classi essenziali non può essere ergodica. Infatti, denotando con C_1 e C_2 due tali classi e con i, j due stati in C_1 e C_2 rispettivamente, abbiamo

$$p^{(n)}(i,j) \ = \ 0 \qquad \text{per ogni} \ \ n \in \mathbb{N}$$

mentre

$$\lim_{n\to+\infty} p^{(n)}(j,j) \ = \ \pi_j^* \ = \ \frac{1}{\sum_{n\geq 1} n f^{(n)}(j,j)} \ \neq 0$$

quindi il limite di $p^{(n)}(u,j)$ per $n \to +\infty$ dipende dallo stato di partenza u.

Corollario 3.16 *Una catena di Markov finita è ergodica se e solo se possiede un'unica classe essenziale e quest'ultima è aperiodica.*

Notiamo infine che la precedente dimostrazione della relazione (3.12) può essere adattata per provare che, se vi sono più componenti essenziali, j è uno stato essenziale aperiodico e i uno stato transiente, allora

$$\lim_{n\to+\infty} p^{(n)}(i,j) \ = \ f(i,j)\pi_j^*$$

In questo caso infatti avremo $f(i,j) \leq 1$ e non è detto che $f(i,j) = 1$, mentre le altre proprietà sono del tutto valide e la formula ottenuta è assolutamente intuitiva.

Appendice A

Richiami di base

In questa appendice ricordiamo le nozioni matematiche di base utilizzate in queste dispense, molte delle quali sono state introdotte in corsi iniziali della laurea triennale e sono generalmente note agli studenti dei corsi di laurea magistrale. I temi principali riguardano il calcolo delle probabilità, l'algebra lineare e alcune proprietà elementari di teoria dei numeri. La presentazione di questi argomenti è necessariamente sintetica e non ha alcuna pretesa di completezza. Vogliamo semplicemente aiutare la comprensione del materiale presentato nei capitoli precedenti. Lo scopo è anche quello di fissare la notazione usata in queste dispense, riassumendo concetti e proprietà fondamentali, solitamente presentati nei libri di testo in maniera più estesa.

A.1 Calcolo delle Probabilità

Ricordiamo innanzitutto che uno *spazio di probabilità* è una tripla $(\Omega, \mathcal{B}, Pr)$ che gode delle seguenti proprietà:

1. Ω è un insieme, detto *spazio campione*, i cui elementi sono chiamati punti campione;

2. $\mathcal{B}$ è una σ-algebra su Ω, ovvero una famiglia di sottoinsiemi di Ω tale che

   - $\Omega \in \mathcal{B}$,
   - per ogni $A \in \mathcal{B}$ il complementare A^c appartiene a $\mathcal{B}$,
   - per ogni sequenza $\{A_i\}_{i\geq 1}$ di elementi di $\mathcal{B}$, abbiamo $\bigcup_{i\geq 1} A_i \in \mathcal{B}$.

   Gli elementi di $\mathcal{B}$ sono chiamati eventi casuali o semplicemente *eventi*;

3. Pr è una misura di probabilità su $\mathcal{B}$, ovvero $Pr : \mathcal{B} \longrightarrow [0,1]$ tale che

   - $Pr(\Omega) = 1$,
   - per ogni sequenza $\{A_i\}_{i\geq 1}$ di elementi disgiunti di $\mathcal{B}$
     $Pr(\bigcup_{i\geq 1} A_i) = \sum_{i\geq 1} Pr(A_i)$ (additività numerabile).

Possiamo facilmente costruire uno spazio di probabilità su un qualunque insieme finito. Sia per esempio Ω un insieme finito e sia $\mathcal{B}$ la famiglia di tutti i sottoinsiemi di Ω. Considera una funzione $p : \Omega \longrightarrow \mathbb{R}$ tale che $0 \leq p(\omega) \leq 1$ per ogni $\omega \in \Omega$ e inoltre $\sum_{\omega\in\Omega} p(\omega) = 1$; definiamo allora Pr come la funzione che ad ogni $A \in \mathcal{B}$ associa il valore $Pr(A) = \sum_{\omega \in A} p(\omega)$. Si verifica facilmente che la tripla $(\Omega, \mathcal{B}, Pr)$ appena definita è uno spazio di probabilità.

Varie altre proprietà possono essere derivate dagli assiomi precedenti. Elenchiamo nel seguito solo le più semplici. Dato uno spazio di probabilità $(\Omega, \mathcal{B}, Pr)$ si verifica che:

- per ogni $A \in \mathcal{B}$, $Pr(A^c) = 1 - Pr(A)$;
- per ogni $A, B \in \mathcal{B}$, se $A \subseteq B$ allora $Pr(A) \leq Pr(B)$;
- se $\{A_1, A_2, \ldots, A_n\}$ è un insieme finito di elementi di $\mathcal{B}$ disgiunti allora
  $Pr(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n Pr(A_i)$ (additività finita).

Inoltre, data una sequenza $\{A_n\}_{n\geq 1}$ di eventi casuali (appartenenti a $\mathcal{B}$), definiamo il limite inferiore e il limite superiore di $\{A_n\}_{n\geq 1}$ mediante le relazioni

$$\liminf_{n\to+\infty} A_n = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k \qquad \limsup_{n\to+\infty} A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k \ .$$

Entrambi tali limiti sono anche loro eventi casuali. Il limite inferiore è l'insieme di tutti gli elementi $\omega \in \Omega$ che appartengono a tutti gli insiemi A_n, $n \geq 1$, tranne al più ad un numero finito di questi; il limite superiore è invece l'insieme di tutti gli elementi $\omega \in \Omega$ che appartengono a infiniti insiemi A_n, $n \geq 1$. Lo loro probabilità è data da

$$Pr(\liminf_{n\to+\infty} A_n) = \lim_{n\to+\infty} Pr(\bigcap_{k\geq n} A_k) \qquad Pr(\limsup_{n\to+\infty} A_n) = \lim_{n\to+\infty} Pr(\bigcup_{k\geq n} A_k)$$

Si dice anche che $\{A_n\}_{n\geq 1}$ converge a un insieme A se i due limiti inferiore e superiore coincidono e il loro valore è proprio A.

Un'altra nozione fondamentale è quella di indipendenza. Dato uno spazio di probabilità $(\Omega, \mathcal{B}, Pr)$, due eventi casuali A, B $(\in \mathcal{B})$ si dicono *indipendenti* se $Pr(A \cap B) = Pr(A)Pr(B)$. Intuitivamente, due eventi sono indipendenti se l'occorrenza di uno dei due non influisce sulla probabilità di occorrenza dell'altro. Più in generale, una famiglia qualsiasi di eventi casuali $\{A_j\}_{j\in I}$ si dice indipendente se, per ogni sottoinsieme finito di indici $S \subseteq I$,

$$Pr(\bigcap_{j\in S} A_j) = \prod_{j\in S} Pr(A_j)$$

Dati due eventi A, B, dove $Pr(A) > 0$, la *probabilità condizionata* di B rispetto ad A, denotata $Pr(B \mid A)$, è definita da

$$Pr(B \mid A) = \frac{Pr(A \cap B)}{Pr(A)}$$

e intuitivamente rappresenta la probabilità di occorrenza di B dato che si è verificato A. Nota che la tripla $(\Omega, \mathcal{B}, Pr_A)$, dove $Pr_A(B) = Pr(B \mid A)$ per ogni $B \in \mathcal{B}$, è uno spazio di probabilità. Se invece $Pr(A) = 0$ si pone per convenzione $Pr(B \mid A) = 0$. Chiaramente, se A e B sono indipendenti allora $Pr(B \mid A) = Pr(B)$ e $Pr(A \mid B) = Pr(A)$. Inoltre, le seguenti proprietà possono essere facilmente dimostrate a partire dalle definizioni e sono spesso utilizzate per calcolare la probabilità di un evento casuale:

- data una famiglia finita di eventi $A_1, A_2, \ldots, A_n \in \mathcal{B}$ tali che $Pr(\bigcap_i A_i) > 0$, abbiamo

  $Pr(\bigcap_{i=1}^n A_i) = Pr(A_1)\ Pr(A_1 \mid A_2)\ Pr(A_3 \mid A_1 \cap A_2) \cdots Pr(A_n \mid A_1 \cap \cdots \cap A_{n-1})$.

- Sia $\{A_j\}_{j\in I}$ una famiglia finita o numerabile di eventi che formano una partizione di Ω. Allora, per ogni $B \in \mathcal{B}$, abbiamo $Pr(B) = \sum_{j\in I} Pr(B \cap A_j)$ e quindi

  $Pr(B) = \sum_{j\in I'} Pr(B \mid A_j) Pr(A_j)$

  dove I' è il sottoinsieme di I degli indici j tali che $Pr(A_j) > 0$.

Una variabile aleatoria X su uno spazio di probabilità $(\Omega, \mathcal{B}, Pr)$ è una funzione $X : \Omega \longrightarrow \mathbb{R}$ tale che, per ogni $a \in \mathbb{R}$, l'insieme $\{\omega \in \Omega \mid X(\omega \leq a\}$ appartiene a $\mathcal{B}$. Nota che in tale definizione X dipende da Ω e da $\mathcal{B}$ ma non dalla misura di probabilità Pr. Si può inoltre provare che se X è variabile aleatoria su $(\Omega, \mathcal{B}, Pr)$ allora, per ogni insieme di Borel [1] $A \subseteq \mathbb{R}$, l'insieme $\{\omega \in \Omega \mid X(\omega) \in A\}$ appartiene a $\mathcal{B}$. Per semplificare la notazione si è soliti rappresentare tali insiemi senza menzionare i punti campione $\omega \in \Omega$; così per esempio, gli insiemi $\{\omega \in \Omega \mid X(\omega) \in A\}$ e $\{\omega \in \Omega \mid X(\omega \leq a\}$ sono denotati rispettivamente dalle espressioni $\{X \in A\}$ e $\{X \leq a\}$, mentre le loro probabilità sono espresse da $Pr(X \in A)$ e $Pr(X \leq a)$.

Ogni variabile aleatoria X su $(\Omega, \mathcal{B}, Pr)$ è dotata di una funzione distribuzione $F_X : \mathbb{R} \longrightarrow [0,1]$, definita da

$$F_X(a) = Pr(X \leq a) \qquad \text{per ogni } a \in \mathbb{R}.$$

È facile verificare che F_X è sempre non decrescente, che $\lim_{a \to -\infty} F_X(a) = 0$, $\lim_{a \to +\infty} F_X(a) = 1$ e inoltre che F_X è una funzione continua dalla destra su tutto $\mathbb{R}$, cioè $\lim_{0<h \to 0} F_X(a+h) = F_X(a)$ per ogni $a \in \mathbb{R}$. Infine si verifica facilmente che $Pr(a < X \leq b) = F_X(b) - F_X(a)$, per ogni $a < b$.

Una variabile aleatoria X su $(\Omega, \mathcal{B}, Pr)$ si dice *discreta* se l'insieme dei suoi valori $X(\Omega)$ è finito o infinitamente numerabile. In questo caso, posto $X(\Omega) = \{a_1, a_2, \ldots\}$, possiamo definire la funzione $p_X : \{a_1, a_2, \ldots\} \longrightarrow [0,1]$, tale che $p_X(a_i) = Pr(X = a_i)$, per ogni i. Otteniamo così $Pr(X \in A) = \sum_{a_i \in A} p_X(a_i)$, per ogni $A \subseteq X(\Omega)$. La funzione p_X è chiamata *funzione probabilità* di X. Si verifica facilmente che, per ogni indice i

$$F_X(a_i) = \sum_{j : a_j \leq a_i} p_X(a_j), \qquad \text{mentre} \qquad p_X(a_j) = \lim_{0<h \to 0} (F_X(a_j) - F_X(a_j - h))$$

permettendo così di calcolare F_X conoscendo i valori di p_X e viceversa.

Un esempio tradizionale è dato dalla variabile aleatoria binomiale X_{np} di parametri n e p, dove $n \in \mathbb{N}$ e $0 \leq p \leq 1$; essa rappresenta il numero di successi in una sequenza di n prove Bernoulliane indipendenti, dove ogni prova ha probabilità di successo p e probabilità di insuccesso $q = 1 - p$. I possibili valori di X_{np} sono $\{0, 1, \ldots, n\}$ e la sua funzione probabilità è data da

$$p_{X_{np}}(k) = \binom{n}{k} p^k q^{n-k}, \qquad \text{per ogni } k = 0, 1, \ldots, n.$$

Un altro esempio è costituito dalla variabile aleatoria geometrica X_p di parametro p, $0 < p \leq 1$. Essa è definita come il numero di prove necessarie per ottenere il primo successo in una sequenza illimitata di prove Bernoulliane indipendenti (dove p rappresenta di nuovo la probabilità di successo). In questo caso, l'insieme dei valori di X_p è quello degli interi positivi, mentre sua funzione di probabilità è definita da $p_{X_p}(k) = (1-p)^{k-1} p$, per ogni $k = 1, 2, \ldots$.

Un variabile aleatoria X su $(\Omega, \mathcal{B}, Pr)$ si dice *continua* se esiste una funzione $f_X : \mathbb{R} \longrightarrow \mathbb{R}$ tale che $f_X(t) \geq 0$ per ogni $t \in \mathbb{R}$, e inoltre $F_X(a) = \int_{-\infty}^{a} f_X(t) dt$ per ogni $a \in \mathbb{R}$. La funzione f_X è detta *funzione densità* di X e coincide con la derivata di F_X. Essa è chiaramente integrabile in $\mathbb{R}$, $\int_a^b f_X(t) dt = Pr(a < X < b)$ per ogni $a < b$ e, più generalmente, $Pr(X \in A) = \int_A f(t) dt$ per ogni insieme di Borel $A \subseteq \mathbb{R}$ (integrale di Lebesgue). Osserviamo infine che, per ogni variabile aleatoria continua X, $Pr(X = a) = 0$ per ogni $a \in \mathbb{R}$.

Un esempio classico è quello della funzione densità normale (o Gaussiana) di parametri $m, s \in \mathbb{R}$, $s > 0$. Essa è definita dalla funzione

$$f(t) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(t-m)^2}{2s^2}}, \qquad -\infty < t < +\infty$$

[1] La famiglia degli insiemi di Borel in $\mathbb{R}$ è definita come la più piccola σ-algebra su $\mathbb{R}$ che contiene gli intervalli della retta reale.

Una variabile aleatoria si dice normale se è la sua funzione densità è normale. Tali variabili aleatorie devono la loro importanza al teorema del limite centrale (che ricordiamo nel seguito) e sono legate alla valutazione degli errori nelle approssimazioni di un valore casuale basate sulla media aritmetica di un insieme di osservazioni.

Un altro esempio è quello della densità esponenziale di parametro $\lambda > 0$, definita dalla funzione

$$f(t) = \begin{cases} 0 & \text{se } t < 0 \\ \lambda e^{-\lambda t} & \text{se } t \geq 0 \end{cases}$$

Anche in questo caso una variabile aleatoria si dice esponenziale se la sua funzione densità è esponenziale. Tali variabili aleatorie godono della seguente notevole proprietà che ne giustifica l'utilizzo come modello per lo studio di tempi di attesa di certi eventi casuali: se X è una variabile aleatoria esponenziale allora $Pr(X > a+b \mid X > a) = Pr(X > b)$ per ogni coppia di valori reali $a, b > 0$. Questo significa intuitivamente che X è priva di memoria; in altre parole, se X rappresenta il tempo di attesa di un dato evento, allora la proprietà precedente asserisce che se abbiamo già aspettato un tempo a la probabilità di attesa di un successivo tempo b è la stessa che si avrebbe all'inizio del processo.

Ricordiamo ora la definizione dei momenti di una variabile aleatoria. Sia g una funzione di variabile reale, definita su tutto $\mathbb{R}$, a valori reali o complessi. Data una variabile aleatoria discreta X, a valori nell'insieme $\{a_1, a_2, \ldots\}$, definiamo il valor medio $E[g(X)]$ di $g(X)$ mediante l'uguaglianza $E[g(X)] = \sum_i g(a_i) p_X(a_i)$, assumendo che la serie sia assolutamente convergente (ovvero che $\sum_i |g(a_i)| p_X(a_i) < +\infty$). Analogamente, data una variabile aleatoria X continua, definiamo il valor medio $E[g(X)]$ di $g(X)$ mediante $E[g(X)] = \int_{-\infty}^{+\infty} g(t) f_X(t) dt$, assumendo anche in questo caso che la funzione sia assolutamente integrabile.

In particolare, se g è la funzione identità, otteniamo il *valor medio* (o semplicemente *media*) della variabile aleatoria X,

$$E[X] = \sum_i a_i p_X(a_i) \qquad \text{oppure} \qquad E[X] = \int_{-\infty}^{+\infty} t f_X(t) dt.$$

Osserviamo che, se X è una variabile aleatoria discreta e assume valori in $\mathbb{N}$, allora

$$E[X] = \sum_{k \geq 1} Pr(X \geq k)$$

Per ogni $m > 0$, $E[X^m]$ è il momento m-esimo di X, rispettivamente

$$E[X^m] = \sum_i a_i^m p_X(a_i) \qquad \text{oppure} \qquad E[X^m] = \int_{-\infty}^{+\infty} t^m f_X(t) dt.$$

Inoltre, il momento m-esimo di $X - E[X]$, ovvero $E[(X - E[X])^m]$, è chiamato momento centrale m-esimo di X. Di particolare importanza è il momento centrale secondo di X, denotato $var(X)$ e solitamente chiamato *varianza* di X. Si verifica facilmente che

$$var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

La varianza di X è intuitivamente una misura della dispersione di X intorno al suo valor medio. Vale inoltre la nota disuguaglianza di Chebyshev

$$Pr(|X - E[X]| \geq a) \leq \frac{Var(X)}{a^2} \qquad \text{per ogni } a > 0$$

che consente di valutare attraverso la varianza la probabilità di distanza di X da $E[X]$. Un'altra disuguaglianza utile è quella di Markov che vale però solo per le variabili aleatorie positive: Se Y è una variabile casuale che assume solo valori in $\mathbb{R}_+$ allora, per ogni reale $a > 0$, abbiamo

$$Pr(Y \geq a) \leq \frac{E[Y]}{a}$$

Altre proprietà derivabili dalle definizioni sono le seguenti (nelle quali c è una costante qualsiasi):

1. $E[c] = c$;
2. $E[cX] = cE[X]$;
3. $var(cX) = c^2 var(X)$;
4. $E[cg(X)] = cE[g(X)]$;
5. $E[c_1 g_1(X) + c_2 g_2(X)] = c_1 E[g_1(X)] + c_2 E[g_2(X)]$
6. $E[g_1(X)] \leq E[g_2(X)]$ se $g_1(x) \leq g_2(x)$ per ogni $x \in \mathbb{R}$.

Come esempi, ricordiamo che la media di una variabile aleatoria binomiale X_{np} è $E(X_{np}) = np$ mentre $var(X_{np}) = np(1-p)$. Nel caso di una variabile aleatoria geometrica X_p abbiamo $E(X_p) = 1/p$ e $var(X_p) = (1-p)/p^2$. Inoltre, una variabile aleatoria esponenziale di parametro λ possiede media $1/\lambda$ e varianza $1/\lambda^2$, mentre m e s^2 sono rispettivamente la media e la varianza di una variabile aleatoria Gaussiana di parametri m e s.

Ricordiamo infine che la media di una variabile aleatoria non è sempre definita. Questo si verifica quando la serie (o l'integrale) che definisce il valor medio non è convergente. Per esempio, se definiamo $Y = 2^{X_p}$, dove X_p è una variabile aleatoria geometrica di parametro p ($0 < p < 1$), allora abbiamo

$$E[Y] = E[2^{X_p}] = \sum_{n \geq 1} 2^n (1-p)^{n-1} p = 2p \sum_{n \geq 0} (2(1-p))^n$$

Quest'ultima serie è convergente solo nel caso $1/2 < p < 1$; quindi, per ogni $0 < p \leq 1/2$ il valor medio di Y non è definito.

L'operatore E permette infine di definire la funzione caratteristica $\varphi_X(t)$ di una variabile aleatoria X. Essa è definita come la funzione $\varphi_X : \mathbb{R} \to \mathbb{C}$, che associa ad ogni $t \in \mathbb{R}$ il valore $\varphi_X(t) = E[e^{itX}]$, dove i è l'unità immaginaria, $i = \sqrt{-1}$. Poiché $|e^{itx}| = 1$, la funzione $\varphi_X(\cdot)$ è ben definita su tutto $\mathbb{R}$, per qualunque variabile aleatoria X. Inoltre $\varphi_X(0) = 1$, $|\varphi_X(t)| \leq 1$ per ogni $t \in \mathbb{R}$, e $\varphi_X(\cdot)$ è (uniformemente) continua su tutto $\mathbb{R}$. Come esempio ricordiamo che se X è una variabile aleatoria Gaussiana di media 0 e varianza 1, allora la sua funzione caratteristica è data da $\varphi_X(t) = e^{-t^2/2}$.

La proprietà fondamentale di $\varphi_X(\cdot)$, per una generica variabile aleatoria X, è che essa caratterizza la funzione distribuzione di X. In altre parole, due variabili aleatorie hanno la stessa funzione distribuzione se e solo se possiedono la stessa funzione caratteristica. Inoltre, essa consente di determinare la distribuzione limite di sequenze di variabili aleatorie: data una sequenza di variabili aleatorie $\{X_n\}_{n \geq 1}$, ciascuna delle quali possiede funzione distribuzione F_n e funzione caratteristica φ_n, e una variabile aleatoria Y con funzione distribuzione G e funzione caratteristica ψ, si verifica che

$$\lim_{n \to +\infty} F_n(x) = G(x) \quad \text{per tutti i punti } x \in \mathbb{R} \text{ di continuità di } F$$

se e solo se

$$\lim_{n\to+\infty} \varphi_n(t) = \psi(t) \quad \text{per ogni } t \in \mathbb{R}.$$

Discutiamo ora la nozione di indipendenza per variabili aleatorie. Innanzitutto osserviamo che se X e Y sono variabili aleatorie definite sul medesimo spazio di probabilità, anche $X+Y$ è una variabile aleatoria definita su tale spazio, essa associa ad ogni punto campione ω il valore $X(\omega)+Y(\omega)$. Per lo stesso motivo, $X-Y$, XY, X/Y (assumendo $Y \neq 0$) sono variabili aleatorie. In particolare si verifica che

$$E[X+Y] = E[X]+E[Y],$$

mentre

$$var(X+Y) = var(X)+var(Y)+2cov(X,Y),$$

dove $cov(X,Y)$ è la covarianza di X e Y, definita da $cov(X,Y) = E\left[(X-E[X])(Y-E[Y])\right]$, che rappresenta intuitivamente il grado di correlazione tra X e Y.

Detto questo, consideriamo n variabili aleatorie $X_1, X_2, \ldots, X_n$ definite sullo stesso spazio di probabilità; tali variabili sono dette *indipendenti* se per ogni n-pla $B_1, \ldots, B_n$ di insiemi di Borel in $\mathbb{R}$, abbiamo

$$Pr(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} Pr(X_i \in B_i)$$

In particolare, se le variabili aleatorie X_i sono discrete, la condizione precedente equivale a porre

$$Pr(X_1 = a_1, \ldots, X_n = a_n) = \prod_{i=1}^{n} Pr(X_i = a_i)$$

per ogni n-pla di valori reali $a_1, \ldots, a_n$. La definizione si estende alle sequenze infinite di variabili aleatorie $\{X_i\}_{i\geq 1}$ dicendo che tali variabili sono indipendenti se, per ogni n, le $X_1, X_2, \ldots, X_n$ sono variabili aleatorie indipendenti.

Tra le proprietà più semplici ricordiamo che se le variabili aleatorie X e Y sono indipendenti allora $E[XY] = E[X]E[Y]$ e quindi $cov(X,Y) = 0$, il che implica $var(X+Y) = var(X)+var(Y)$.

In generale possiamo dire che lo studio del comportamento di sequenze di variabili aleatorie indipendenti è tra gli argomenti principali del tradizionale calcolo delle probabilità. Ricordiamo qui in una forma relativamente semplice due proprietà fondamentali, ovvero la legge dei grandi numeri e il teorema del limite centrale. Esse riguardano le somme parziali di una sequenza di variabili aleatorie. Intuitivamente, la prima stabilisce che, con alta probabilità, tale somma si concentra intorno al suo valor medio; la seconda mostra che la funzione distribuzione di tale somma approssima una distribuzione normale. Più precisamente, supponiamo che $\{X_i\}_{i\geq 1}$ sia una sequenza di variabili aleatorie indipendenti aventi la stessa distribuzione. In tale ipotesi valgono le seguenti proprietà:

1. (Legge dei grandi numeri) Se m è il valor medio di X_1 (e quindi di tutte le $X_i, i \geq 1$) allora, con probabilità 1,

   $$\lim_{n\to+\infty} \frac{X_1+X_2+\cdots+X_n}{n} = m$$

   Questo implica in particolare una proprietà più debole, ovvero che per ogni $\varepsilon > 0$,

   $$\lim_{n\to+\infty} Pr\left\{\left|\frac{X_1+X_2+\cdots+X_n}{n} - m\right| > \varepsilon\right\} = 0.$$

2. (Teorema del limite centrale) Se m e s sono la media e la varianza di X_1 (e quindi di tutte le $X_i, i \geq 1$) allora, per ogni $x \in \mathbb{R}$

$$\lim_{n \to +\infty} Pr \left\{ \frac{X_1 + \cdots + X_n - nm}{\sqrt{sn}} \leq x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

   Lo strumento tradizionale utilizzato per provare questa proprietà è quello delle funzioni caratteristiche: si fa cioé vedere che la funzione caratteristica della somma parziale (opportunamente normalizzata) converge in ogni punto alla funzione caratteristica della normale standard (quella di media 0 e varianza 1).

Un tema di studio e di ricerca naturale in questo contesto è quello di determinare le ipotesi che consentono di estendere il teorema del limite centrale o la legge dei grandi numeri al caso di sequenze di variabili aleatorie dipendenti o non ugualmente distribuite. Il più semplice esempio di variabili aleatorie dipendenti è proprio fornito dalle catene di Markov e alcune funzioni di tali quantità hanno proprio una distribuzione limite normale.

A.2 Matrici e vettori

Dati due interi positivi r e s, una matrice di dimensione $r \times s$ è una famiglia di $r \cdot s$ valori a_{ij}, dove $i \in \{1, 2, \ldots, r\}$ e $j \in \{1, 2, \ldots, s\}$, che noi rappresentiamo come tavola bidimensionale nella forma

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1s} \\ a_{21} & a_{22} & \cdots & a_{2s} \\ \cdots & \cdots & & \cdots \\ a_{r1} & a_{r2} & \cdots & a_{rs} \end{pmatrix}$$

oppure mediante l'espressione $[a_{ij}]_{1 \leq i \leq r,\ 1 \leq j \leq s}$ o ancora, più semplicemente, da $[a_{ij}]$ quando la dimensione $r \times s$ è nota. Il valore r è il numero di righe di A mentre s è il numero di colonne. Gli elementi a_{ij} sono detti coefficienti (o componenti) della matrice. Se $r = s$ diremo che la matrice è quadrata e che r è il suo ordine. Un vettore riga è una matrice di dimensione $1 \times r$ mentre un vettore colonna è una matrice di dimensione $r \times 1$; in entrambi i casi diremo che r è la dimensione del vettore. Un vettore a di dimensione r è generalmente considerato come vettore colonna e rappresentato nella forma

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_r \end{pmatrix}$$

Dato un insieme T denotiamo con $T^{r \times s}$ la famiglia di tutte le matrici di dimensione $r \times s$ a coefficienti in T. Analogamente, con T^r denoteremo l'insieme di tutti i vettori (colonna) di dimensione r a coefficienti in T. Così, $\mathbb{R}^{3 \times 7}$ è l'insieme di tutte le matrici a coefficienti reali di dimensione 3×7, mentre $\mathbb{C}^4$ è l'insieme di tutti vettori di dimensione 4 a coefficienti complessi.

La matrice trasposta di una matrice $A = [a_{ij}]$ di dimensione $r \times s$ è la matrice $[b_{ji}]$ di dimensione $s \times r$, tale che $b_{ji} = a_{ij}$ per ogni $1 \leq i \leq r,\ 1 \leq j \leq s$; la matrice trasposta di A si ottiene quindi scambiando le righe e le colonne e viene solitamente rappresentata da A'. Così, se a è un vettore colonna, il suo trasposto a' è il corrispondente vettore riga. Per esempio, posto

$$A = \begin{pmatrix} 3 & 7 & 1 & 5 \\ 2 & 3 & 1 & 0 \\ 1 & -3 & 4 & -1 \end{pmatrix} , \qquad a = \begin{pmatrix} 2 \\ -3 \\ 6 \end{pmatrix}$$

allora

$$A' = \begin{pmatrix} 3 & 2 & 1 \\ 7 & 3 & -3 \\ 1 & 1 & 4 \\ 5 & 0 & -1 \end{pmatrix} , \qquad a' = (2,\ -3,\ 6) \ .$$

Una sottomatrice di una matrice A si ottiene cancellando da A alcune righe e/o alcune colonne.

Ricordiamo ora le tradizionali relazioni ed operazioni tra matrici. D'ora in poi consideriamo solo matrici e vettori a coefficienti reali o complessi. Due matrici $A = [a_{ij}]$ e $B = [b_{ij}]$ di dimensione $r \times s$ si dicono uguali, in simboli $A = B$, se $a_{ij} = b_{ij}$ per ogni coppia di indici i e j ($1 \leq i \leq r$, $1 \leq j \leq s$). Analogamente, $A < B$ significa che A e B sono a coefficienti reali e che $a_{ij} < b_{ij}$ per ogni $i = 1, \ldots, r$ e ogni $j = 1, \ldots, s$. In modo del tutto simile si definiscono le relazioni $A \leq B$, $A \geq B$, $A > B$. Inoltre una matrice reale $A = [a_{ij}]$ si dice positiva, in simboli $A > 0$, se $a_{ij} > 0$ per ogni i e ogni j; A si dice invece non negativa, $A \geq 0$, se $a_{ij} \geq 0$ per tutte le coppie di indici.

Una matrice quadrata A si dice *simmetrica* se coincide con la sua trasposta, ovvero $A = A'$. Inoltre una matrice quadrata $[a_{ij}]$ si dice *diagonale* se tutti i suoi coefficienti sono nulli tranne gli elementi a_{ii}. Tali elementi formano la cosiddetta diagonale principale della matrice. Chiaramente, tutte le matrici diagonali sono simmetriche. Tra queste ricordiamo la matrice identità di ordine r, solitamente denotata con I_r, definita ponendo $I_r = [\delta_{ij}]$, dove i valori δ_{ij} sono i coefficienti di Kronecker, definiti dalle equazioni $\delta_{ii} = 1$ e $\delta_{ij} = 0$ per ogni $i \neq j$. Quando l'ordine r è sottinteso, I_r sarà più semplicemente rappresentata con I, omettendo l'indice r.

Altre due matrici spesso usate sono la matrice $0_{r\times s}$ di dimensione $r \times s$, le cui componenti sono tutte 0 e la matrice $E_{r\times s}$ della stessa dimensione, i cui coefficienti sono tutti 1. I corrispondenti vettori (colonna) saranno rappresentati da 0_r e e_r. Anche in questo caso gli indici saranno omessi quando le dimensioni sono sottointese.

Tre sono le operazioni principali tra matrici a coefficienti complessi: il prodotto per uno scalare, la somma e la moltiplicazione. Il prodotto cA di una matrice $A = [a_{ij}]$ per una costante $c \in \mathbb{C}$ è la matrice $cA = [ca_{ij}]$. In particolare, abbiamo $(-1)A = -A = [-a_{ij}]$, $0A = [0]$. La somma $A + B$ di due matrici $A = [a_{ij}]$ e $B = [b_{ij}]$ della stessa dimensione è data da $A + B = [a_{ij} + b_{ij}]$. La somma di matrici è un'operazione associativa e commutativa; inoltre il prodotto per una costante è distributivo rispetto alla somma, ovvero $c(A+B) = cA + cB$ e $(c+d)A = cA + dA$.

Il prodotto di due matrici $A = [a_{ij}]$ e $B = [b_{ij}]$ è definito quando il numero di colonne di A è uguale al numero di righe di B; se $r \times s$ è la dimensione di A e $s \times t$ quella di B allora il prodotto AB è la matrice $AB = [c_{ij}]$ di dimensione $r \times t$, tale che $c_{ij} = \sum_{k=1}^{s} a_{ik} b_{kj}$ per tutti gli indici i, j.

Il prodotto di matrici è associativo ma non commutativo; tale operazione è distributiva rispetto alla somma di matrici, ovvero $(A+B)C = AC + BC$ e $A(B+C) = AB + AC$. Per ogni matrice A di dimensione $r \times s$ valgono anche le seguenti identità: $I_r A = A I_s = A$, $0_{t\times r} A = 0_{t\times s}$, $A 0_{s\times t} = 0_{r\times t}$. Inoltre la trasposta di un prodotto è sempre il prodotto inverso delle trasposte, ovvero $(AB)' = B'A'$. Infine, per ogni $n \in \mathbb{N}$, la potenza A^n di una matrice quadrata A è il prodotto di A per sè stessa n volte, con la convenzione che $A^0 = I$. Il prodotto di potenze di una stessa matrice è commutativo.

Il determinante di una matrice quadrata $A = [a_{ij}]$ di ordine r è definito da

$$\det(A) = \sum_{*} \pm a_{1j_1} a_{2j_2} \cdots a_{rj_r}$$

dove la somma è estesa a tutte le permutazioni $(j_1, j_2, \ldots, j_r)$ di $(1, 2, \ldots, r)$ e il segno $\pm$ è positivo se la permutazione è pari, negativo se è dispari. Se $\det(A) = 0$ diciamo che A è singolare e questo avviene se e solo se una delle sue righe (o colonne) si può esprimere come combinazione lineare delle altre. Viceversa, se $\det(A) \neq 0$ diremo che A è regolare o non singolare e in questo caso le righe (e le colonne) di A sono linearmente indipendenti. Il determinante può essere calcolato usando i cofattori; il cofattore $\mathrm{cof}(A)_{ij}$ è il valore

$$\mathrm{cof}(A)_{ij} = (-1)^{i+j} \det(A_{ij}) ,$$

dove A_{ij} è la sottomatrice che si ottiene cancellando da A la riga i-esima e la colonna j-esima. Si può provare che per ogni $1 \leq i,\ j \leq r$,

$$\det(A) = \sum_{k=1}^{r} a_{ik} \mathrm{cof}(A)_{ik} = \sum_{k=1}^{r} a_{kj} \mathrm{cof}(A)_{kj} .$$

Inoltre l'aggiunto di A, denotato $\mathrm{adj}(A)$, è la trasposta della matrice dei cofattori, ovvero

$$\mathrm{adj}(A) = [\mathrm{cof}(A)_{ij}]' . \tag{A.1}$$

Infine, se A e B sono matrici quadrate dello stesso ordine, allora $\det(AB) = \det(A)\det(B)$.

L'inversa di una matrice quadrata A è la matrice A^{-1} tale che $AA^{-1} = A^{-1}A = I$. Si può dimostrare che A^{-1} esiste se e solo se A non è singolare, ovvero $\det(A) \neq 1$, nel qual caso abbiamo

$$A^{-1} = \frac{1}{\det(A)} \mathrm{adj}(A)$$

Chiaramente $I_r^{-1} = I_r$ per ogni intero $r > 0$ e si può provare che, se A e B sono matrici quadrate non singolari dello stesso ordine, allora $(AB)^{-1} = B^{-1}A^{-1}$ e $(A')^{-1} = (A^{-1})'$.

Molte proprietà delle catene di Markov sono basate sullo studio delle potenze di matrici. Per questo motivo in questa sezione ricordiamo alcune caratteristiche di base di tali quantità.

Proposizione A.1 *Se A è una matrice quadrata tale che $A^n \to 0$ (matrice nulla) per $n \to +\infty$, allora $(I - A)^{-1}$ esiste e inoltre*

$$(I - A)^{-1} = I + A + A^2 + \cdots + A^n + \cdots$$

Dimostrazione. È facile verificare che per ogni intero $n > 0$

$$I - A^n = (I - A)(I + A + A^2 + \cdots + A^{n-1})$$

Sappiamo inoltre che il determinante della matrice a sinistra dell'uguaglianza tende a 1; quindi per ogni n abbastanza grande $\det(I - A^n) \neq 0$. Allora anche il determinante della matrice a destra dell'uguaglianza è diverso da 0 e, poiché il determinante di un prodotto è il prodotto dei determinanti, otteniamo $\det(I - A) \neq 0$. Questo implica che $(I - A)^{-1}$ è ben definita. Se ora moltiplichiamo l'equazione precedente per $(I - A)^{-1}$ otteniamo

$$(I - A)^{-1}(I - A^n) = I + A + A^2 + \cdots + A^{n-1}$$

e facendo tendere n all'infinito si ricava il risultato. □

A.3 Autovalori e autovettori

Data una matrice quadrata $A = [a_{ij}]$ di ordine r, diciamo che un numero complesso $\lambda \in \mathbb{C}$ è un *autovalore* di A se esiste un vettore $u \neq 0_r$ tale che
$$u'A = \lambda u' .$$
In questo caso si dice che u è *autovettore sinistro* corrispondente a λ. L'insieme S_λ ottenuto considerando tutti gli autovettori sinistri di A corrispondenti a λ e aggiungendo il vettore 0_r, forma uno spazio vettoriale, sottospazio di $\mathbb{C}^r$; S_λ è anche chiamato autospazio sinistro di A corrispondente a λ e la sua dimensione è detta *molteplicità geometrica* di λ.

Gli autovalori di A possono anche essere definiti come le radici (nella variabile x) dell'equazione
$$\det(xI_r - A) = 0$$
chiamata *equazione caratteristica* di A. Si mostra facilmente che le due definizioni sono equivalenti. Inoltre si chiama *molteplicità algebrica* di λ la sua molteplicità come radice del polinomio $\det(xI_r - A)$. Poiché l'equazione caratteristica ha grado r ci possono essere al più r autovalori distinti di A e in questo caso tutti hanno molteplicità algebrica 1. Si può dimostrare che la molteplicità geometrica di un autovalore è sempre minore o uguale alla sua molteplicità algebrica. Chiaramente, se la molteplicità algebrica di un autovalore è 1 anche quella geometrica è 1. Il viceversa non è sempre vero.

Così come abbiamo definito l'autospazio sinistro, possiamo definire l'autospazio destro. Dato un autovalore λ di A chiamiamo *autovettore destro* corrispondente a λ un vettore $v \neq 0_r$ di dimensione r, tale che $Av = \lambda v$. L'insieme di tutti questi vettori con l'aggiunta di 0_r forma uno spazio vettoriale D_λ, chiamato autospazio destro corrispondente a λ. Si verifica che la sua dimensione coincide con quella di S_λ e quindi con la molteplicità geometrica di λ.

Si può anche provare che se $\lambda_1, \lambda_2, \ldots, \lambda_t$ sono autovalori distinti di A allora ogni t-pla di autovettori sinistri $u_1, \ldots, u_t$, dove $u_i \in S_{\lambda_i}$ per ogni i, forma un insieme di vettori indipendenti. Una proprietà analoga vale per gli autovettori destri. Questo implica che il massimo numero di autovettori sinistri (destri) indipendenti è minore o uguale alla somma delle molteplicità geometriche degli autovalori; a sua volta tale somma può essere minore di r e questo capita quando un autovalore possiede molteplicità geometrica minore di quella algebrica.

Un modo per calcolare facilmente la potenza A^n di una matrice A passa attraverso la sua diagonalizzazione. Una matrice quadrata A si dice *diagonalizzabile* se esiste una matrice non singolare U e una matrice diagonale D tali che
$$D = UAU^{-1} .$$
In questo caso i coefficienti che si trovano sulla diagonale principale di D sono gli autovalori di A, cioé $D = [\lambda_i \delta_{ij}]$, dove i λ_i sono gli autovalori di A. Inoltre le righe di U e le colonne di U^{-1} sono rispettivamente autovettori sinistri e autovettori destri corrispondenti agli autovalori. Questo implica che una matrice è diagonalizzabile se e solo se le molteplicità geometriche di tutti gli autovalori coincidono con quelle algebriche. Chiaramente abbiamo $A^n = U^{-1}D^nU$ dove D^n è la matrice diagonale data da $D^n = [\lambda_i^n \delta_{ij}]$.

Due sono i casi di matrici diagonalizzabili particolarmente importanti: il caso delle matrici reali simmetriche e quello delle matrici con autovalori distinti. Se A è una matrice reale simmetrica allora valgono le seguenti proprietà: 1) tutti gli autovalori di A sono reali; 2) gli autovettori destri e sinistri di uno stesso autovalore coincidono; 3) A è diagonalizzabile, cioè $A = U^{-1}DU$ con D e U definiti come sopra e 4) si può scegliere la matrice degli autovettori U in modo che sia una matrice ortogonale, ovvero $U^{-1} = U'$ (i.e., A possiede r autovettori tra loro ortogonali).

Se invece A è una matrice quadrata di ordine r che possiede r autovalori distinti $\lambda_1, \dots, \lambda_r$, allora è facile provare che per ogni $i = 1, \dots, r$ esistono un autovettore sinistro u_i e uno destro v_i, entrambi corrispondenti a λ_i, tali che $u_i' v_i = 1$ e $u_i' v_j = 0$, per ogni $i \neq j$. Possiamo allora considerare la matrice U che si ottiene accostando le r righe $u_1', \dots, u_r'$ e la matrice V ottenuta accostando le colonne $v_1, \dots v_r$. Tali matrici, entrambe quadrate di ordine r, sono non singolari e inoltre abbiamo $UV = I$, cioè $U = V^{-1}$ e $V = U^{-1}$. Esse consentono di diagonalizzare la matrice A; dallo loro stessa definizione infatti si ricava $UA = DU$ e $AV = VD$, dove $D = [\lambda_i \delta_{ij}]$ e quindi otteniamo

$$A = U^{-1} DU = VDU \ .$$

Quest'ultima relazione può essere riscritta nella forma

$$A = \sum_{i=1}^{r} \lambda_i v_i u_i'$$

che solitamente è chiamata *rappresentazione spettrale* di A e deve la sua importanza alla immediata deduzione della equazione

$$A^n = \sum_{i=1}^{r} \lambda_i^n v_i u_i' \tag{A.2}$$

valida per ogni intero $n > 0$, che consente di calcolare facilmente le potenze di A.

A.4 Massimo comun divisore

In questa sede ricordiamo alcune proprietà del massimo comun divisore con particolare riferimento alle famiglie di interi chiuse rispetto alla somma. L'obiettivo principale è quello di provare che i sottoinsiemi di $\mathbb{N}$ chiusi rispetto alla somma contengono tutti i multipli positivi del loro massimo comun divisore eccetto al più un numero finito di questi. Tale proprietà, qui enunciata nel teorema A.4, è utile in particolare nell'analisi del periodo delle matrici irriducibili sviluppata nel capitolo 1. La presentazione è qui ripresa da [9].

Sia S un insieme di numeri interi, eventualmente infinito, contenente almeno un elemento a diverso da 0. Il massimo comun divisore di S può essere definito nel modo tradizionale: consideriamo la famiglia dei divisori comuni a tutti gli elementi di S; essa non è vuota perché contiene il numero 1, ed è certamente finita perchè il numero di divisori di a è finito. Possiamo quindi considerare il massimo valore in tale famiglia, detto appunto *massimo comun divisore* di S.

Ricordiamo infine che un insieme $S \subseteq \mathbb{Z}$ è chiuso rispetto alla somma se, per ogni $a, b \in S$, anche l'intero $a + b$ appartiene a S. Chiaramente un tale insieme contiene infiniti elementi se possiede almeno un intero diverso da 0. Definizione analoghe si possono dare per la sottrazione e le altre operazioni aritmetiche.

Lemma A.2 *Sia $S \subseteq \mathbb{Z}$ un insieme chiuso rispetto alla somma e alla sottrazione, contenente un intero diverso da* 0*. Allora S possiede un minimo elemento positivo m e inoltre S coincide con l'insieme dei multipli di m, ovvero $S = \{cm \mid c \in \mathbb{Z}\}$.*

Dimostrazione. Se $a \neq 0$ appartiene a S allora anche $0 = a - a$ e $-a = 0 - a$ appartengono a S. Quindi S contiene almeno un elemento positivo e possiamo considerare il minimo m dei suoi elementi positivi, $m = \min\{t \in S \mid t > 0\}$. Per le proprietà di chiusura di S, gli interi $-km$ e km appartengono a S per ogni $k = 1, 2, \dots$. Mostriamo ora che ogni elemento di S è un multiplo di m. Infatti, se $t \in S$ allora, dividendo

t per m, determiniamo due interi q e r, $0 \leq r < m$ tali che $t = qm + r$; di conseguenza anche $r = t - qm$ appartiene a S e quindi $r = 0$ perché m è il minimo elemento positivo in S ed r deve essere minore di m. Ne segue che t è un multiplo di m. □

Lemma A.3 *Il massimo comun divisore d di un insieme finito di numeri interi $\{b_1, b_2, \ldots, b_k\}$ è combinazione lineare a coefficienti interi di $b_1, b_2, \ldots, b_k$, ovvero esistono k coefficienti $c_1, c_2, \ldots, c_k \in \mathbb{Z}$ tali che $d = \sum_{j=1}^{k} c_j b_j$.*

Dimostrazione. Sia S l'insieme delle combinazioni lineari $\sum_{j=1}^{k} i_j b_j$ al variare degli interi i_j in $\mathbb{Z}$. Poiché S è chiuso rispetto alla somma e alla sottrazione, per il Lemma A.2, esiste un elemento positivo minimo $m \in S$ tale che ogni elemento di S è multiplo di m. Mostriamo che $d = m$. Infatti, si verifica che $0 < d \leq m$, poiché d, essendo divisore di ogni b_j, è divisore di ogni intero in S. D'altra parte, poiché ogni b_j appartiene a S, m divide ogni intero b_j per cui $m \leq d$. Quindi $m = d$ e questo prova che anche d è combinazione lineare dei b_j. □

Teorema A.4 *Ogni insieme non vuoto di interi positivi chiuso rispetto alla somma contiene tutti i multipli positivi del suo massimo comun divisore eccetto al più un numero finito di questi.*

Dimostrazione. Sia U un insieme non vuoto di interi positivi chiuso rispetto alla somma e sia d il suo massimo comun divisore. Proviamo il teorema supponendo $d = 1$. Se infatti fosse $d > 1$ basterebbe dividere tutti gli elementi di U per d per ricondursi al caso precedente. Allora, $d = 1$ è anche il massimo comun divisore di un numero finito di elementi di U, che denotiamo $b_1, b_2, \ldots, b_k$; quindi, per il Lemma A.3, esistono k interi $i_1, i_2, \ldots, i_k \in \mathbb{Z}$ tali che

$$1 = \sum_{j=1}^{k} i_j b_j = u - v \tag{A.3}$$

dove u è la somma degli elementi positivi nella sommatoria e $-v$ la somma di quelli negativi. Se $v = 0$ allora, nella sommatoria precedente, tutti gli elementi sono positivi o nulli, il che implica che qualche $b_j \in U$ è uguale a 1, ed essendo U chiuso rispetto alla somma U deve contenere tutti gli interi positivi. Se invece $v > 0$ allora abbiamo $u, v \in U$ perché somme di elementi di U. Consideriamo allora un qualunque intero $n \geq v^2$ e mostriamo che $n \in U$. Infatti, dividendo n per v otteniamo due interi q e r tali che $n = qv + r$, $0 \leq r < v$ e $v \leq q$. Per la (A.3) abbiamo $n = qv + r(u - v) = ru + (q - r)v$, ed essendo $q > r$, otteniamo che n è somma di elementi di U e di conseguenza appartiene a U. □

Appendice B

Dimostrazione del teorema di Perron-Frobenius

In questo capitolo presentiamo una dimostrazione del teorema di Perron-Frobenius per le matrici primitive. L'enunciato preciso che qui riportiamo per comodità è stato dato nel teorema 1.7. Esso afferma che ogni matrice primitiva possiede un unico autovalore di modulo massimo e fornisce varie altre proprietà riguardanti i suoi autovalori e autovettori. Questi risultati cosentono di determinare l'espressione asintotica delle potenze di matrici primitive che illustriamo nell'ultima sezione e che è legata alle proprietà ergodiche delle catene di Markov primitive. La presentazione che qui sviluppiamo è sostanzialmente tratta da [9].

Teorema B.1 *Sia $A \geq 0$ una matrice primitiva. Allora esiste un autovalore λ di A che gode delle seguenti proprietà:*

1. *λ è reale e maggiore di 0;*
2. *ogni autovalore μ di A diverso da λ verifica la relazione $|\mu| < \lambda$;*
3. *λ ammette autovettori destri e sinistri strettamente positivi;*
4. *λ è radice semplice dell'equazione caratteristica di A (la sua molteplicità algebrica è 1).*

La costante λ è chiamata *autovalore di Perron-Frobenius* della matrice A, oppure *autovalore principale* di A. La sua costruzione è basata sulle proprietà della funzioni continue: data la matrice $A \in \mathbb{R}_+^{m\times m}$, l'idea è quella di definire una opportuna funzione a valori reali sull'insieme $\mathbb{R}_+^m$ che risulta continua sulla iper-sfera $S \subseteq \mathbb{R}_+^m$ di raggio unitario: l'autovalore λ sarà proprio il valore massimo assunto dalla funzione su S e l'autovettore sinistro v sarà invece il punto nel quale tale valore viene ottenuto.

B.1 Costruzione dell'autovalore principale

Iniziamo illustrando il passo principale del teorema, ovvero la costruzione di λ.

Sia $A \in \mathbb{R}_+^{m\times m}$ una matrice primitiva, $A = [a_{ij}]$. Considera la funzione

$$r : \mathbb{R}_+^m \longrightarrow \mathbb{R}_+ \cup \{+\infty\}$$

tale che, per ogni $x \in \mathbb{R}^m_+$,

$$r(x) = \min_j \left\{ \frac{(x'A)_j}{x_j} \right\} = \min_j \left\{ \frac{\sum_i x_i a_{ij}}{x_j} \right\}$$

dove si assume che $\frac{(x'A)_j}{x_j} = +\infty$ se $x_j = 0$.

Tale funzione gode delle seguenti proprietà:

a) per ogni $x \in \mathbb{R}^m_+$, se $x \neq \underline{0}$ allora $0 \leq r(x) < +\infty$;

b) per ogni $x, y \in \mathbb{R}^m_+$ e ogni $c > 0$, se $y = cx$ allora $r(x) = r(y)$. In altre parole, r assume gli stessi valori sulle rette che passano per $\underline{0}$;

c) la funzione r non è continua in $\underline{0}$. Infatti è facile verificare che

$$r(1/k, \ldots, 1/k) = \min_j \left\{ \sum_i a_{ij} \right\} < +\infty = r(\underline{0}) \; ;$$

d) la funzione r assume nel punto $\underline{1} = (1, \ldots, 1)$ il minimo delle somme delle colonne di A, ovvero $r(\underline{1}) = \min_j \left\{ \sum_i a_{ij} \right\}$;

e) per ogni $x \in \mathbb{R}^m_+$, se $x \neq \underline{0}$ dalla definizione di r si deduce

$$x' r(x) \leq x'A \; . \tag{B.1}$$

Inoltre questo implica

$$r(x) \leq \frac{(x'A)\underline{1}}{x'\underline{1}} \leq \max_i \left\{ \sum_j a_{ij} \right\}$$

ovvero, $r(x)$ è sempre minore o uguale al massimo delle somme delle righe di A.

Per la proprietà **b)** i valori finiti di r sono esattamente quelli assunti dalla funzione nella iper-sfera parziale $S = \{x \in \mathbb{R}^m_+ \mid x'x = 1\}$. Possiamo allora definire il valore λ dato da

$$\lambda = \sup_{x \in S} \{r(x)\}$$

Tale valore è ben definito perché per le proprietà **d)** ed **e)** abbiamo

$$\min_j \left\{ \sum_i a_{ij} \right\} \leq \lambda \leq \max_i \left\{ \sum_j a_{ij} \right\}$$

Questo prova anche che $\lambda > 0$ perché A non può possedere una colonna di elementi nulli (altrimenti non sarebbe irriducibile).

Una nota proprietà delle funzioni a valori reali afferma che ogni funzione continua in un sottoinsieme compatto (cioè chiuso e limitato) di $\mathbb{R}^m$ assume in tale insieme un valore massimo. Poichè S è un sottoinsieme compatto di $\mathbb{R}^m$, esiste un elemento $v \in S$ tale che

$$r(v) = \lambda = \sup_{x \in S} \{r(x)\}$$

Dimostriamo ora che λ è proprio un autovalore di A e che v è un suo autovettore sinistro. Considera il vettore $z = v'A - v'r(v)$. Nota che tale vettore è uguale a $\underline{0}$ se e solo se la proprietà è verificata. Osserva inoltre che per la relazione (B.1) abbiamo $v'r(v) \leq v'A$ e quindi

$$z' = v'A - v'r(v) \geq \underline{0}$$

Supponi ora che $z \neq \underline{0}$ e considera un intero $k > 0$ tale che $A^k > 0$ (tale intero esiste sempre per la primitività di A). Allora $z'A^k > \underline{0}$ e quindi

$$(v'A^k)A - (v'A^k)r(v) > \underline{0}$$

ovvero, per ogni indice i, $((v'A^k)A)_i > (v'A^k)_i r(v)$ il che implica

$$r(v'A^k) > r(v)$$

Questo significa che la funzione r assume nel punto $v'A^k$ un valore maggiore di $r(v) = \lambda$ e questo è assurdo per la stessa definizione di λ. Di conseguenza $z' = v'A - v'r(v) = \underline{0}$ e quindi

$$v'A = \lambda v'$$

Abbiamo così provato il punto 1 del teorema 1.7.

Massimalità di λ

Proviamo ora il punto 2. Sia $\mu \in \mathbb{C}$ un autovalore di A diverso da λ. Dimostriamo innanzitutto che $|\mu| \leq \lambda$. A tal fine considera un autovettore $x \in \mathbb{C}^m$ corrispondente a μ ($x \neq \underline{0}$):

$$\mu x' = x'A \tag{B.2}$$

Passando ai moduli otteniamo

$$|\mu x_j| \leq \sum_i |x_i| a_{ij} \qquad (\forall j)$$

Assumendo ora la solita convenzione nel caso $x_j = 0$, possiamo scrivere

$$|\mu| \leq \frac{\sum_i |x_i| a_{ij}}{|x_j|} \qquad (\forall j)$$

Questo implica

$$|\mu| \leq \min_j \left\{ \frac{\sum_i |x_i| a_{ij}}{|x_j|} \right\} = r(|x|) \leq \lambda$$

poiché λ è il massimo valore che la funzione r può assumere nei punti $|x| \in \mathbb{R}^m_+$ diversi da $\underline{0}$.

Proviamo ora che se $|\mu| = \lambda$ allora $\mu = \lambda$. Infatti

$$\lambda \cdot |x_j| = |\mu x_j| = \left| \sum_i x_i a_{ij} \right| \leq \sum_i |x_i| a_{ij} \tag{B.3}$$

$$\lambda \cdot |x_j| \leq \sum_i |x_i| a_{ij} \qquad (\forall j) \tag{B.4}$$

Di nuovo, essendo λ il massimo di r in S, possiamo applicare il ragionamento svolto sopra per provare che il vettore z è nullo. Otteniamo così $\lambda \cdot |x_j| = \sum_i |x_i| a_{ij}$ per tutti gli indici j. Allora anche la disuguaglianza in (B.3) è una uguaglianza:

$$\left| \sum_i x_i a_{ij} \right| = \sum_i |x_i| a_{ij}$$

Tale relazione asserisce che il modulo di una somma di numeri complessi è uguale alla somma dei moduli. Questo è vero solo se tutti i numeri hanno lo stesso esponente, ovvero $x_j = \rho_j e^{i\theta}$ per qualche $0 \leq \theta < 2\pi$ (con $\rho_j \geq 0$). Così, dividendo i termini della equazione (B.2) per $e^{i\theta}$, si ottiene

$$\mu \rho_j = \sum_i \rho_i a_{ij} \qquad (\forall j)$$

che implica $\mu \in \mathbb{R}_+$ e quindi $\mu = \lambda$.

B.2 Molteplicità geometrica dell'autovalore principale

Dimostriamo ora che λ possiede autovettori positivi, ovvero il punto 3 del teorema di Perron-Frobenius. Questa proprietà consente di provare la molteplicità geometrica di λ, insieme ad un'altra proprietà riguardante gli autovalori delle matrici ottenute da A riducendo uno o più dei suoi coefficienti. Quest'ultima, che sarà utilizzata spesso in seguito, afferma che riducendo il valore di qualunque entrata della matrice primitiva si riduce anche il modulo dell'autovalore massimo.

Tali proprietà sono le seguenti:

i) l'autovettore v definito nella sezione precedente è strettamente positivo;

ii) l'autovalore λ possiede molteplicità geometrica 1;

iii) se B è una matrice tale che $0 \leq B \leq A$ e β è un autovalore di B allora $|\beta| \leq \lambda$; inoltre, se $|\beta| = \lambda$ allora $B = A$.

Proposizione B.2 *L'autovettore v definito nella sezione precedente è strettamente positivo.*

Dimostrazione. Per la primitività di A esiste un intero $k > 0$ tale che $A^k > 0$; quindi dall'equazione $v'A = \lambda v'$ si ottiene facilmente

$$v'A^k = \lambda^k v'$$

Osserva ora che $v \geq \underline{0}$ poiché v appartiene all'insieme S definito nella sezione precedente. Di conseguenza $v'A^k > \underline{0}$ e l'equazione precedente implica $v > \underline{0}$. □

Proposizione B.3 *L'autovalore λ possiede molteplicità geometrica* 1.

Dimostrazione. Considera un autovettore sinistro x di λ e proviamo che $x = cv$ per qualche $c \in \mathbb{C}$. Osserva che anche il vettore $x_+ = |x|$ è autovettore corrispondente a λ e, ripetendo il ragionamento svolto nella proposizione B.2, otteniamo $x_+ > \underline{0}$. Allora, per ogni $c \neq 0$, anche $\eta = v - cx$ è autovettore corrispondente a λ, così come lo è il vettore $\eta_+ = |\eta|$ e anche per lui vale $\eta_+ > \underline{0}$. Ora, se x non è multiplo di v, possiamo scegliere c in modo tale che $\eta \neq \underline{0}$ e che qualche componente η_i sia uguale a 0. Questo implicherebbe $\eta_{+_i} = 0$, il che è assurdo perché $\eta_+ > \underline{0}$. Ne segue che x deve essere multiplo di v.

I ragionamenti svolti finora si possono ripetere per gli autovettori destri invece che per quelli sinistri. Tutte le corrispondentti proprietà si dimostrano in modo simmetrico. Inoltre l'unicità dell'autovalore principale, dimostrata nella sezione precedente, garantisce che i due valori di λ ottenuti sono uguali. □

Proposizione B.4 *Se B è una matrice tale che $0 \leq B \leq A$ e β è un autovalore di B allora $|\beta| \leq \lambda$; inoltre, se $|\beta| = \lambda$ allora $B = A$.*

Dimostrazione. Sia y un autovettore destro di B corrispondente all'autovalore β: $By = \beta y$. Allora

$$|\beta y| = |By| \leq B|y| \leq A|y|$$

che implica

$$|\beta| \cdot |y| \leq A\,|y| \tag{B.5}$$

Moltiplicando ora per v' si ottiene

$$|\beta| \cdot v'|y| \leq v'A|y| = \lambda v'|y|$$

e dividendo per $v'|y|$ si ricava

$$|\beta| \leq \lambda$$

che prova la prima parte dell'enunciato.

Supponi ora che $|\beta| = \lambda$; dalla equazione (B.5) abbiamo

$$\lambda|y| \leq A|y|$$

che, per il solito ragionamento sulla massimalità di λ, implica $\lambda|y| = A|y| > \underline{0}$. Di conseguenza, si ricava la seguente relazione

$$A\,|y| = \lambda|y| = |\beta| \cdot |y| = |B\,y| \leq B\,|y| \leq A\,|y|$$

cioè $B|y| = A|y|$ con $|y| > \underline{0}$, ed essendo $0 \leq B \leq A$, questo è possibile solo se $B = A$. □

B.3 Molteplicità algebrica dell'autovalore principale

Questa sezione è dedicata allo studio della molteplicità algebrica di λ. Qui proviamo che λ è radice semplice del polinomio caratteristico di A e che inoltre la matrice $\mathrm{adj}(\lambda I - A)$ è strettamente positiva (la definizione di $\mathrm{adj}(\cdot)$ è data in (A.1)). Nota che la matrice $(\lambda I - A)$ non ammette inversa perché λ è autovalore di A. Tuttavia $\mathrm{adj}(\lambda I - A)$ è sempre ben definita e come vedremo gioca un ruolo importante nell'analisi delle potenze di A.

Proposizione B.5 *La matrice adj$(\lambda I - A)$ è positiva.*

Dimostrazione. Innazitutto osserviamo che per ogni $c \in \mathbb{C}$ valgono le seguenti uguaglianze:

$$(cI - A)\,\mathrm{adj}(\lambda I - A) = \det(\lambda I - A) \cdot I \tag{B.6}$$

$$\mathrm{adj}(\lambda I - A)\,(cI - A) = \det(\lambda I - A) \cdot I \tag{B.7}$$

Infatti tali relazioni sono ovvie per tutti i c diversi dagli autovalori di A. Se c è un autovalore di A le equazioni valgono per continuità.

Consideriamo allora la matrice $\mathrm{adj}(\lambda I - A)$. Per le due equazioni precedenti, ogni sua riga x' e ogni sua colonna y soddisfano

$$x'(\lambda I - A) = \underline{0} \qquad \text{e} \qquad (\lambda I - A)y = \underline{0}$$

Questo significa che per ogni riga x' ci sono due possibilità : $x = \underline{0}$ oppure x' è autovettore sinistro di A corrispondente a λ e quindi $x > \underline{0}$. Lo stesso vale per ogni colonna y. Di conseguenza una proprietà analoga vale per la matrice $\mathrm{adj}(\lambda I - A)$: tutti i suoi elementi sono maggiori di 0, oppure sono tutti uguali a 0. Proviamo allora che un suo elemento è positivo (e quindi lo sono tutti). Considera l'elemento di indice $1,1$, dato da $\det(\lambda I_{m-1} - A^{11})$, dove A^{11} è ottenuto da A cancellando la prima riga e la prima colonna. Posto $B = \begin{bmatrix} 0 & \underline{0}' \\ \underline{0} & T^{11} \end{bmatrix}$, vale $0 \leq B \leq A$ e quindi per la proposizione B.4 λ non può essere autovalore di B. Questo significa che

$$0 \neq \det(\lambda I - B) = \lambda \cdot \det(\lambda I_{m-1} - A^{11})$$

e quindi anche $\det(\lambda I_{m-1} - A^{11})$ è diverso da 0. Di conseguenza $\mathrm{adj}(\lambda I - A) > 0$. □

Proposizione B.6 *La molteplicità algebrica di λ è 1.*

Dimostrazione. Considera la funzione $\varphi(z) = \det(zI - A)$. Derivando i termini dell'equazione (B.6) si ottiene

$$\mathrm{adj}(zI - A) + (zI - A)\frac{d}{dz}\mathrm{adj}(zI - A) = \varphi'(z)I$$

Moltiplicando a sinistra per v' e ponendo $z = \lambda$ si ottiene

$$v'\mathrm{adj}(\lambda I - A) = \varphi'(\lambda)v' \tag{B.8}$$

Per le proprietà dimostrate in precedenza sappiamo che il vettore di sinistra è a componenti positive. Di conseguenza abbiamo anche $\varphi'(\lambda) > 0$ e questo significa che λ è radice semplice dell'equazione caratteristica di A. Nota che la (B.8) prova che $\varphi'(\lambda)$ è autovalore di $\mathrm{adj}(\lambda I - A)$ e v un suo autovettore sinistro. □

Il seguente corollario è una conseguenza delle due proposizioni precedenti e fornisce una espressione precisa della matrice $\mathrm{adj}(\lambda I - A)$ in funzione degli autovettori corrispondenti a λ.

Corollario B.7 *Se v' e u sono rispettivamente autovettore sinistro e autovettore destro di A rispetto a λ, tali che $v'u = 1$, allora*

$$\frac{adj(\lambda I - A)}{\varphi'(\lambda)} = uv'$$

Dimostrazione. Nella dimostrazione della proposizione B.5 abbiamo visto che ogni riga di $\mathrm{adj}(\lambda I - A)$ è un autovettore sinistro corrispondente a λ e quindi della forma cv' per qualche costante c, perché la molteplicità geometrica di λ è 1. Per la positività di $\mathrm{adj}(\lambda I - A)$, tali costanti sono tutte positive e quindi la matrice si può scrivere nella forma

$$\mathrm{adj}(\lambda I - A) = \begin{bmatrix} c_1 v' \\ c_2 v' \\ \cdots \\ c_m v \end{bmatrix}$$

dove $c = (c_1, c_2, \ldots, c_m) \in \mathbb{R}^m_+$. Ne segue che anche c è autovettore destro di A rispetto a λ e di conseguenza esiste una costante $a > 0$ tale che

$$\mathrm{adj}(\lambda I - A) = a \cdot uv'$$

Inoltre, moltiplicando a sinistra per v' si ottiene

$$v'\mathrm{adj}(\lambda I - A) = av'$$

e per la (B.8) si ricava $a = \varphi'(\lambda)$. □

B.4 Potenze di matrici primitive

Una delle principali conseguenze del teorema di Perron Frobenius riguarda l'ordine di grandezza delle potenze di matrici primitive. Essa asserisce che la potenza A^n di una matrice primitiva A cresce con ordine di grandezza pari a λ^n, dove λ è l'autovalore principale di A.

Prima di presentare questo risultato e la relativa prova, ricordiamo la seguente proprietà delle funzioni razionali: per ogni costante $a > 0$ e ogni intero $m > 0$, la funzione $(1-az)^{-m}$ ammette uno sviluppo in serie di Taylor con centro in 0, valido per ogni $z \in \mathbb{C}$ tale che $|z| < 1/a$, della forma seguente:

$$(1-az)^{-m} = \sum_{n=0}^{+\infty} \binom{m+n-1}{n} a^n z^n \qquad \text{e inoltre} \qquad \binom{m+n-1}{n} a^n = \Theta(n^{m-1}a^n) \tag{B.9}$$

Teorema B.8 *Sia $A \in \mathbb{R}_+^{m\times m}$ una matrice primitiva e sia λ il suo autovalore principale. Allora, al crescere di n, abbiamo*

$$A^n = \lambda^n uv' + O(\rho^n) \qquad (|\rho| < \lambda) \tag{B.10}$$

dove v' e u sono rispettivamente autovettore sinistro e autovettore destro di A corrispondenti a λ, normalizzati in modo tale che $v'u = 1$.

Dimostrazione. Siano $\mu_1, \ldots, \mu_j$ gli autovalori di A diversi da λ e siano $m_1, \ldots, m_j$ le rispettive molteplicità. Allora vale la relazione

$$\det(I - Az) = (1-\lambda z)(1-\mu_1 z)^{m_1} \cdots (1-\mu_j z)^{m_j}$$

Per il teorema 1.7, $|\mu_i| < \lambda$ per ogni i e quindi la funzione (matriciale) $(I-Az)^{-1} = \frac{\mathrm{adj}(I-Az)}{\det(I-Az)}$ è ben definita e regolare per ogni $|z| < 1/\lambda$. Inoltre si può provare che nello stesso intorno $|z| < 1/\lambda$ la funzione $(I-Az)^{-1}$ ammette lo sviluppo di Taylor con centro in 0

$$(I-Az)^{-1} = \sum_{n=0}^{+\infty} A^n z^n$$

Usando la tradizionale decomposizione delle funzioni razionali in fratti semplici, possiamo allora rappresentare la serie precedente nella forma

$$\sum_{n=0}^{+\infty} A^n z^n = (I-Az)^{-1} = \frac{\mathrm{adj}(I-Az)}{\det(I-Az)} = \tag{B.11}$$

$$\frac{X}{(1-\lambda z)} + \frac{Y_1(z)}{(1-\mu_1 z)^{m_1}} + \cdots \frac{Y_j(z)}{(1-\mu_j z)^{m_j}} \tag{B.12}$$

dove $X \in \mathbb{R}_+^{m\times m}$ mentre le $Y_i(z)$ sono matrici di polinomi in z. Applicando ora la (B.9) e ricordando che $|\mu_i| < \lambda$ per ogni i, l'equazione precedente implica

$$A^n = X\lambda^n + \mathrm{O}(\rho^n)$$

per qualche $|\rho| < \lambda$.

Valutiamo ora la matrice X. Considera di nuovo l'uguaglianza

$$(I-Az)^{-1} = \frac{X}{(1-\lambda z)} + \frac{Y_1(z)}{(1-\mu_1 z)^{m_1}} + \cdots \frac{Y_j(z)}{(1-\mu_j z)^{m_j}}$$

valida per ogni $|z| < 1/\lambda$. Moltiplicando entrambi i termini per $(1-\lambda z)$ e facendo tendere z a $1/\lambda$ da sinistra sulla retta reale, si ottiene

$$\lim_{z \to 1/\lambda^-} (1-\lambda z)(I-Az)^{-1} = X \tag{B.13}$$

Sempre per $z \to 1/\lambda^-$, il termine di sinistra è asintotico a

$$(1-\lambda z)\,\frac{\operatorname{adj}(I-Az)}{\det(I-Az)} \sim (1-\lambda z)\,\frac{z^{m-1}\operatorname{adj}(I\frac{1}{z}-A)}{z^m \det(I\frac{1}{z}-A)} \sim \tag{B.14}$$

$$\sim \lambda \operatorname{adj}(I\lambda - A)\frac{(1-\lambda z)}{\det(I\frac{1}{z}-A)} \sim \frac{\operatorname{adj}(I\lambda-A)}{\varphi'(\lambda)} \tag{B.15}$$

dove $\varphi(x) = \det(Ix - A)$ e l'ultimo passaggio è ottenuto applicando la regola di de l'Hopital alla frazione precedente. Sostituendo l'espressione ottenuta nel limite (B.13) si deduce

$$X = \frac{\operatorname{adj}(I\lambda - A)}{\varphi'(\lambda)}$$

e il risultato segue allora dal corollario B.7. $\square$

Bibliografia

[1] P. Brémaud, *Markov Chains: Gibbs fields, Monte Carlo Simulation and Queues*, Springer, New Jork, 1998.

[2] W. Feller, *An Introduction to Probability Theory and Applications*, J. Wiley, 1968.

[3] O. Häggström, *Finite Markov Chains and Algorithmic Applications*, London Mathematical Society, 2003.

[4] M. Iosifescu, *Finite Markov Processes and Their Applications*, J. Wiley and Sons, 1980.

[5] D. Isaacson, R. Madsen, *Markov Chains Theory and Applications*, R.E. Krieger Publishing Company, 1985.

[6] J.G. Kemeny e J.L. Snell, *Finite Markov Chains*, Van Nostrand, 1960.

[7] A. Paz, *Introduction to Probabilistic Automata*, Academic Press, 1971.

[8] A. Salomaa e M. Soittola, *Automata-Theoretic Aspects of Formal Power Series*, Springer-Verlag, 1978.

[9] E. Seneta, *Non-negative Matrices and Markov Chains*, Springer–Verlag, 1981.

[10] G. Strang, *Linear Algebra and its Applications*, Harcourt Brace & Company, 1988.

[11] W. Woess, *Catene di Markov e teoria del potenziale nel discreto*, Quaderni dell'Unione Matematica Italiana, n.41, Pitagora Editrice 1996.